Thesis for the Degree of Doctor of Philosophy

Network Analysis of Maritime English Corpus with Multi-word Compounds: Keyword Networks and Collocation Networks

by

Sung-Min Lee

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in the Department
of
English Language and Literature

Korea Maritime and Ocean University

February 2016

Approval

Name: Sung-Min Lee
Degree: Doctor of Philosophy
Title of thesis: Network Analysis of Maritime English Corpus with Multi-word Compounds: Keyword Networks and Collocation Networks

Examining Committee:

Chair:

Dr. Kim, Jae-Hoon
Professor of Computer and Information Engineering
Korea Maritime and Ocean University

Dr. Jhang, Se-Eun
Senior Supervisor, Professor of English Language and Literature
Korea Maritime and Ocean University

Dr. Kwon, Heok-Seung
External Examiner, Professor of English Language and Literature
Seoul National University

Dr. Lee, Soo-Sang
External Examiner, Professor of Library, Archives Information Studies
Pusan National University

Dr. Scott, Mike
External Examiner, Visiting Researcher of Corpus Linguistics
School of Language and Social Sciences
Aston University

Date Approved: December 17, 2015

Abstract

As an official language within the international maritime community, maritime English is one of the branches of English for Specific Purposes (ESP). However, corpus linguists have paid little attention to maritime English. This thesis has two aims. The first aim is to compile a four million word maritime English corpus (MEC) consisting of academy, news, laws, and textbooks. The MEC contains tagged multi-word compounds, which can be called specific purpose terms in maritime English. Tagging multi-word compounds is essential for the ESP study because maritime vocabulary includes a great variety of n-grams such as *ballast water, fore peak bulkhead, container freight station charges,* etc. The second aim is to provide a further explanation of corpus linguistic data, adopting language network analysis and comparing keyword networks with collocation networks.

My idea converging on corpus linguistics and language networks has been originally traced back to researches published by Jones in 1971 and Scott and Tribble in 2006. Jones discussed four types of links between keyword nodes such as strings, stars, cliques, and clumps in her keyword retrieval study. Based on Jones' work, Scott and Tribble hypothesized that keywords could be redrawn as a network of connections to show a picture of understanding about a text or texts. By incorporating corpus linguistics and language networks, this thesis tries to explore what the structures of keywords networks and collocation networks can tell us about maritime English through centrality and cohesion algorithms.

This thesis makes an attempt to answer the following two research questions. First, how can we build a corpus of maritime English to represent specific purpose terms such as multi-word compounds? Second, if language network analysis can be one of the explanatory analyses to make up for the present corpus linguistic descriptions, what can keyword networks and collocation networks tell us about the MEC? In pursuit of my research questions, I review previous studies about the concepts of keyness, collocations, and language networks. I then discuss how to compile the MEC focusing on representativeness, balance, size, and sampling, proposing a method of tagging English multi-word compounds. In addition, I propose a language network analysis in order to give a further explanatory power to the descriptions of maritime English. I compare keyword networks with collocation networks with regard to network structures using centrality and cohesion for the better understanding of maritime English.

In conclusion, my network analysis and critical evaluation led us to clarify and confirm that centrality structures created by eigenvector and betweenness in collocation networks have more advantages over keyword network structures to find general purpose terms. On the other hand, the cohesion community structures created by eigenvector and betweenness in keyword networks distinguish a group of the specific purpose terms from a group of general purpose terms. More specifically, the eigenvector centrality structures in collocation networks represented better results than betweenness centrality in identifying general purpose terms. On the other hand, the eigenvector cohesion community structures in keyword networks represented better results than betweenness in identifying specific purpose terms.

There are several implications of corpus linguistics and pedagogy. My proposed method of tagging English multi-word compounds can help traditional concordance tools and new corpus visualization tools to be developed to a new horizon beyond the current lexico-grammar research because a corpus tagged with multi-word compounds can reflect authentic language usages. In addition, ESP teachers have the advantage to decide what to teach in ESP vocabulary by using network analysis which I have proposed in this thesis.

Acknowledgment

I would like to thank everyone who contributed to the completion of my thesis. First of all, I would like to acknowledge the great support from all my committee members, senior supervisor Se-Eun Jhang, committee chair Jae-Hoon Kim, external examiner Soo-Sang Lee, Heok-Seung Kwon, and Mike Scott. Especially, Dr. Scott reviewed my original ideas, gave valuable comments to my academic paper, and recommended me as a candidate for Global Ph.D. Fellowship. I am grateful to all my faculty members in the department of English language and literature. I have learned a lot of knowledge from them. In addition, I would like to acknowledge Korean Register of Shipping (KR). KR allowed me to access maritime law documents so that I can compile a maritime law corpus which is a part of my maritime English corpus.

I express my deep appreciation to professor Se-Eun Jhang for his invaluable comments and suggestions. He has been directing my research when I began studying at the Korea Maritime and Ocean University. He encouraged me to attend important workshops held by Association of Corpus Linguistics and to present my ideas at academic conferences held by both domestic and international linguistic associations.

I would like to thank National Research Foundation of Korea which granted me Global Ph.D. Fellowship (Project name: *A New Approach to Visualizing a Specialized Corpus Using Network Analysis*, NRF-2013H1A2A1034655) for three years. Without this research fund, my research would not have been successfully completed. Through the Global Ph.D. Fellowship program, I met

professor Pann-Ghill Suh from Ulsan National Institute of Science and Technology, who taught me researchers' attitude. I also met a lot of my fellow graduate students who are majoring different fields. They inspired me to nurture my scientific knowledge.

I would like to express my deep gratitude to my family. My wife, Ae-ran Cho, has endured many years taking care of my little daughter, Seo-Yoon Lee without my assistance. Whenever I was discouraged, she encouraged me and provided every support for me. Without her sacrifices, my thesis would never be completed. I recognize the sacrifices of my mother, Bok-Ja Cho, who is a professional calligrapher and a good painter. Her passion for learning taught me how I have to live as a scholar. I am also grateful to my brother in law, Sung-Eun Kim, and my sister, Seung-Ah Lee for their financial support for me to complete masters' degree in education at The State University of New York in the United States.

Lastly, I would like to thank my father, Jong-Hoon Lee, who is now in heaven. I remember his love and patience toward me. He always taught me to be a pioneer in my field.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1. Introduction

1.1 Focus of Inquiry

As an official language within the international maritime community, maritime English refers to the English language which is used at sea and in port by mariners and is one of the branches of English for Specific Purposes (ESP) (Bocanegra-Valle, 2012). However, there has been too little research in maritime English corpus (MEC) based on both corpus linguistics and language networks. Without the MEC, it is difficult to research authentic language uses of maritime-related academic texts, industry news, and official regulations released by the International Maritime Organization (IMO).

This thesis has two aims. The first aim is to compile a four million word MEC consisting of academy, news, laws, and textbooks genres. The MEC contains the tagged multi-word compounds, which can be called maritime English terms. Tagging multi-word compounds is essential for the ESP study because maritime English vocabulary items comprise a great variety of specific purpose n-gram terms such as *ballast water, fore peak bulkhead, container freight station charge,* etc. The second aim is to provide a further explanatory power to corpus linguistic analysis. By using keyword networks and collocation networks, I analyze the structures created by centrality

measures and cohesion measures in terms of eigenvector and betweenness algorithms. In addition, I conduct a critical evaluation to clarify and to confirm my findings.

My idea converging on corpus linguistics and language networks has been originally traced back to the study of Jones (1971) and Scott and Tribble (2006). Jones (1971: 56) discussed four types of links between keyword nodes such as strings, stars, cliques, and clumps in her keyword retrieval study. Based on Jones' work, Scott and Tribble (2006) hypothesized that keywords could be redrawn as a network of connections by showing Jones' formal pattern of keyword linkage. With regard to language network structure analysis, I compare keyword networks with collocation networks. The keyword networks consist of keywords and their linked keywords within a specified window span, representing textual features considered as text-dependent (Scott, 2010: 43). On the other hand, collocation networks consist of keywords and their co-occurred collocates within a specified window span. Collocation networks deal with linguistic features because the notion of collocation contributes to shaping the meaning of words (Firth, 1957) and also forming lexico-grammar (Halliday, 1961).

This thesis makes an attempt to answer the following two research questions. First, how can we build a corpus of maritime English to represent specific purpose terms such as multi-word compounds? Second, if language network analysis can be one of the explanatory analyses to make up for the present corpus linguistic descriptions, what can keyword networks and collocation networks tell us about the MEC?

1.2 Outline of the Thesis

In pursuit of my research questions, this thesis is organized as follows. Chapter 1 introduces the focus of inquiries and outlines the thesis in terms of how the idea of incorporating corpus linguistics and language networks is studied.

Chapter 2 introduces maritime English, the concepts of keyness and collocations in terms of their the previous studies, the statistical issues, and some problems in order to explore some characteristics of maritime English as ESP in Chapter 3. It also introduces visualization in corpus linguistics including text visualizations and collocation networks. I also review basic concepts for language network construction, previous studies, definitions, and the types of language network constructions in order to provide an explanatory power to corpus data in Chapter 4.

Chapter 3 discusses how to compile the MEC focusing on representativeness, balance, size, and sampling in order to provide answers to the first research question. I propose a new method of tagging English multi-word compounds, holding a debate on the comparison of with and without multi-word compounds with regard to STTR, word lists, n-gram lists, keyword lists, dispersion plots, and GraphColl 1.0. Finally, I summarize the chapter and discuss the implications of the new methods.

Chapter 4 develops my convergence idea incorporating corpus linguistics and language networks in order to provide answers to my second research question. It proposes language network analysis in order to give a further explanatory power to the corpus descriptions by comparing keyword

networks with collocation networks through centrality and cohesion algorithms for the better understanding of maritime English. I create a term speciality algorithm to explain the results of cohesion analysis. I demonstrate that the structure of collocation networks has the advantage for finding general purpose terms and the structure of keyword networks has the advantage for classifying specific purpose term group and general purpose term group. Finally, I summarize the chapter and discuss implications.

Chapter 5 summarizes my findings of the thesis and points out limitations and implications for this study.

Chapter 2. Literature Review

2.1 Introduction

This chapter explores some characteristics of maritime English as ESP and previous studies about keyness and collocations in corpus linguistics. The concepts of keywords in text are discussed along with two strategies for deciding a reference corpus, two statistical measures for keyword analysis, and problems of previous keyword analysis. The concept of a collocation and its recent issues for extraction methods are discussed. This chapter also introduces previous studies of visualization in corpus linguistics. Finally, I will review previous studies of language network analysis with some basic concepts for network analysis, definitions for network concepts, and language network construction and analysis in order to provide an explanatory power for descriptions of maritime English to follow.

2.2 Maritime English as English for Specific Purposes

Maritime English is defined as an official language within the international maritime community, contributing to the safety of navigation and the facilitation of seaborne trade (Trenker, 2000; IMO, 2009; Bocanegra-Valle, 2012). Hence, maritime English as an official language is adopted by IMO[1)]

1) The IMO consists of five major committees such as Maritime Safety Committee,

which is a specialized UN agency located in London, United Kingdom. The IMO has shown interest in training maritime English for mariners because it is one of the most important aspects for the international maritime community and maritime industry.

Maritime English terms are difficult to understand due to the fact that almost all the domain-specific purpose terms are only used in maritime industry. Here are some examples of such specific words and phrases: *ashore, astern, bosun, embark, disembark, funnel, gangway, heave, off-shore, starboard, tamper, wharf, wheelhouse, windlass, tender boat, a port of call, ballast water, flag of convenience,* etc. Thus maritime English is one of the branches of ESP.

Maritime English is inevitable to most people engaging in the maritime industry especially for prospective mariners attending maritime universities and training institutes.[2] Thus, maritime English teaching and learning have been voiced by organizations such as IMO, International Maritime Lectures Association, and International Association of Maritime Universities.

2.3 Keywords in Text

In this section, I discuss keywords with regard to some strategies for deciding a reference corpus, statistical measures for keyword analysis, and problems of previous keyword analysis. In addition, I explore collocations

---

Marine Environment Protection Committee, Legal Committee, Facilitation Committee, and Technical Cooperation Committee.

2) Notably, one of the materials available is Peter Van Kluijven's "The International Maritime Language Programme" (IMLP). The other is the "MarEng project" which provides a web-based maritime English learning tool.

with regard to their concepts and statistical issues.

2.3.1 Strategies for a Reference Corpus

The keyword analysis enables us to find statistically significant positive and negative words by comparing frequencies between a study corpus and a reference corpus (Scott, 1991; 1997; 2001; 2015). With regard to the decision of a reference corpus, there are two strategies of corpus comparison used in corpus linguistics.

The first strategy is a comparison of a study corpus with a larger normative corpus or a general standard corpus such as British National Corpus (BNC) or Corpus of Contemporary American English (COCA). The reference corpus must be a much larger corpus than a study corpus for this comparison. There are at least two kinds of comparisons in this strategy. First, there is a comparison between different types of corpora. For example, a specialized corpus is compared with a general corpus such as BNC or COCA. Second, there is a comparison between the same types of corpora, choosing a text or a group of texts, or a genre out of all genres as a study corpus (Scott and Tribble, 2006). For example, Culpeper (2002) selected six main characters as a study corpus and used the collection of all the other characters' speech in Shakespear's "Romeo and Juliet" as a reference corpus. Scott and Tribble (2006) extracted keywords from "Romeo and Juliet" compared with the other Shakespeare's plays. This method can be called cross-validation[3] (Bengio and Grandvalet, 2004) which has been used for

3) In k-fold cross-validation, the original sample is randomly divided into a k equal

machine learning in the discipline of computational linguistics.

The second strategy is a comparison of two roughly equal-sized corpora. There are at least two types of comparison in this strategy. First, there is a comparison between the different types of corpora. For example, a million word specialized corpus can be compared with a general corpus such as the Brown corpus or the LOB corpus. Second, there is a comparison between the same types of corpora. For instance, there is a comparison of one million word American English corpus such as the Brown corpus with one million word British English corpus such as the LOB corpus in order to examine the differences and similarities between American English and British English.

2.3.2 Statistical Measures for Keyword Analysis

There are a few statistical debates of keyword analysis to argue for a log-likelihood test over a chi-square test. As a statistical point of view, Dunning (1993) discussed that the likelihood ratio could lead to much better statistical results for small corpora.[4] Similarly, Lindquist (2009) argued that

---

sized sample and then a single sample is compared with the remaining k−1 samples. Following k-fold cross validation, this paper extracts keywords by adopting the second method for choosing a reference corpus. Since the MEC has four different genres such as academy, news, laws, and textbooks, each genre can be compared with all the other genres to see what is typical of maritime English genre styles.

4) Dunning (1993: 65) noted that "[I]nstead they use the asymptotic distribution of the generalized likelihood ratio. For text analysis and similar problems, the use of likelihood ratios leads to very much improved statistical results. The practical effect of this improvement is that statistical textual analysis can be done effectively with very much smaller volumes of text than is necessary for

chi-square is not a reliable statistical measure for small samples.[5] Recently, Bertels and Speelman (2013: 24) indicated that log-likelihood could be an efficient statistical measure for both small and large corpora. This can be supported by some journal articles relating to keyword analysis. In "International Journal of Corpus Linguistics", there are five articles regarding keyword analysis (Rayson, 2008; Culpeper, 2009; Fisher-Starcke, 2009; Bednarek, 2012; Bertels and Speelman, 2013) adopting the log-likelihood test to identify keywords, whereas there are not any articles regarding keyword analysis adopting the chi-square test. In "English for Specific Purposes", there are also two articles used the log-likelihood test for finding keywords (Grabowsky, 2015; Le and Harrionton, 2015). With regard to ESP articles, most of the previous ESP studies used log-likelihood instead of other statistical methods (Bertels and Speelman, 2013; Jhang and Lee, 2013b; Grabowsky, 2015; Le and Harrington, 2015).

Keyword analysis enables us to find out some characteristics of a text or a set of texts grouped as a genre including language teaching, forensic linguistics, stylistics, content analysis, and text retrieval, as discussed by Scott (2015). Following Scott (1997), the concept of a keyword means "a word which occurs with unusual frequency in a given text." Keyness is computed by using log-likelihood test (Rayson and Garside, 2000), as shown

---

conventional tests based on assumed normal distributions, and it allows comparisons to be made between the significance of the occurrences of both rare and common phenomenon."

5) Lindquist (2009: 40) noted that "[I]t should be stressed that the chi-square test is not reliable for very small samples, and that with very large samples it is easy to get significant figures. ... it is often possible to judge the significance with the naked eye anyway."

in (1).

(1) $$E_i = N_i \frac{\sum_i O_i}{\sum_i N_i} \quad LL = 2*((a*\log(\frac{a}{E_1})) + (b*\log(\frac{b}{E_2})))$$

Expected value (E_i) uses the number of words in corpus N_i, and an observed frequency (O_i). The calculation of log-likelihood uses an expected value E_i, the frequency of word of a in corpus one, and the frequency of word of b in corpus two. Technically speaking, the higher the keyness value is, the more significant the difference between two frequency scores is. Using Wordsmith tools, keyness in a genre of maritime law is obtained, as shown in Table 2.1.

Table 2.1 Keyness in a genre of maritime law

N	Keywords	Freq.	%	*RC Freq.	RC %	Keyness
1	SHALL	16,380	1.69	1,730	0.06	35,662.63
2	OR	16,497	1.70	11,040	0.37	15,518.80
3	BE	17,869	1.84	18,179	0.61	10,497.82
4	PARAGRAPH	2,138	0.22	268	0.00	4,464.96
5	CONVENTION	3,449	0.36	1,641	0.06	4,200.24
6	PROVISIONS	2,509	0.26	665	0.02	4,152.81
7	ANY	4,220	0.44	2,905	0.10	3,843.97
8	ACCORDANCE	1,941	0.20	302	0.01	3,840.35
9	PROVIDED	2,321	0.24	861	0.03	3,278.91
10	CERTIFICATE	1,519	0.16	176	0.00	3,227.68
.	.	.	.	.	.	.
1,976	SEDIMENT	29	0.00	22	0.00	23.97
1,977	BRING	51	0.00	310	0.01	-24.19
.	.	.	.	.	.	.
2,761	WAS	365	0.04	9,829	0.33	-3,440.39

*RC is a reference corpus

As seen in the above table, there are 2,761 keywords extracted from a genre of maritime law, consisting of 1,976 positive keywords and 785 negative keywords.[6] This keyword list will be used for obtaining linked

6) This keyness method can also be applied to keyclusters extraction and key semantic domain analysis. Clusters (Scott, 2015) are called many different terms indicating multi-word sequences such as pre-fabricated patterns (Hakuta, 1974), collocational frameworks (Renouf, 1991), lexical bundles (Biber et al., 1999), formulaic sequences (Wray, 2000), chunks (O'Keefee, McCarthy, and Carter, 2007), and phraseology (Meunier and Granger, 2008). Researchers have used different criteria for finding multi-word sequences. However, there are common considerations which include the cut-off frequency and dispersion thresholds. For example, frequency thresholds are between 5 to 40 per million words (Biber et al., 1999; Biber et al., 2004; Hyland, 2008; Demmen, 2009; Chen and Baker, 2010). The dispersion threshold is about at least three to five texts, or 10% of texts (Cortes, 2004; Biber and Barbieri, 2007; Hyland, 2008; Chen and Baker, 2010). Keyclusters have been used to identify language usages in many different fields (Baker, 2006; Mahlberg, 2007; Demmen, 2009; Jhang and Lee, 2012). For key semantic domain analysis, there is a Wmatrix Web interface program

keywords and collocates, which will be discussed in Chapter 4.

2.3.3 Problems of Previous Keyword Analysis

Scott (2015) found the following:

> There can be no guarantee that the 'key' words are 'key' in the sense which you may attach to 'key'. An 'important' word might occur once only in a text. They are merely the words which are outstandingly frequent or infrequent in comparison with the reference corpus. (p. 203)

However, there is no doubt that keyword analysis has provided a lot of insights to language researchers for a couple of decades. There are previous studies about keyword analysis in general corpora and historical corpora (Culpeper, 2002; Scott and Tribble, 2006; Mahlberg, 2007; Archer and Culpeper 2009; Demmen, 2009). For example, Culpeper (2002) explored several keywords to identify the characteristics of six main characters from "Romeo and Juliet" by comparing them with all the other characters. Scott and Tribble (2006) also discussed the concepts of keyness and keywords including positive and negative keyness[7]. He extracted keywords from

---

offering the UCREL semantic analysis system (USAS). The semantic tag-set used by the USAS was originally based on McArthur's (1981) Longman Lexicon of Contemporary English. It has a multi-tier structure with 21 major discourse categories with the possibility of further fine-grained subdivision.

7) Scott (2015) explained in his HELP menu that:
[A] word is said to be key if (a) it occurs in the text at least as many times as the user has specified as a minimum Frequency (b) its frequency in the text when compared with its frequency in a reference corpus is such that the statistical probability as computed by an appropriate procedure is smaller than or equal to a p-value specified by the user. A word which is positively key occurs more often than would be expected by chance in comparison with the reference corpus. It is over-used. A word which is negatively key occurs less often than would be

"Romeo and Juliet", and compared the play with other Shakespeare's plays. Mahlberg (2007) discussed keywords to explore some characteristics of the literature of Charles Dickens by comparing a 4.5 million word corpus of Dickens' novels with nineteenth-century writers' corpus. Her study illustrated Dickens' stylistic patterns and additional details about text features of Dickens' novels. Archer and Culpeper (2009) carried out qualitative and quantitative analysis from their socio-pragmatic corpus by using keyness analysis in order to find out style markers which reveal personal pronouns, interjections, imperative verbs, politeness formulae, etc. Demmen (2009) explored keywords from the female characters in Shakespeare's plays. The dialogues of male and female characters were studied to present a variety of language use between male and female characters by using WordSmith tools.

Stubbs (2010) discussed three concepts of keywords regarding different views of keywords. The first sense of keywords is about words and culture deriving from cultural studies, citing Williams (1976)[8] where words such as *work, career, job, labour, (un)employment* are studied. The second sense contains words and texts deriving from comparative quantitative corpus analysis (Scott and Tribble, 2006).[9] The third sense means phrases and

---

expected by chance in comparison with the reference corpus. It is under-used. (pp. 201-202)

8) There are some characteristics of Williams' keywords: (1) intuitively identified words; (2) Widespread words (e.g., *country, expert, family*, and *genius*) and intellectual words (e.g., *alienation, dialectic, hegemony*, and *utilitarian*); (3) Conceptual category words. However, Williams' work did not intend to be a linguistic analysis. The study used a small set of specialized words.

9) Stubbs (2010: 25) stated that "[K]eywords are words which are significantly more frequent in a sample of text than would be expected, given their frequency in a large general reference corpus."

schemas deriving from lexico-grammar, citing Francis (1993) which identified culturally significant units of meaning.[10] Among these three senses of keywords, this thesis is concerned with the second sense.

2.4 Collocations in Text

In this section, I discuss two types of collocations such as significant collocation and non-significant collocation. I also discuss important concepts such as a node, a span, and a collocate. Finally I examine some problems of previous collocation analysis.

2.4.1 Types of Collocations

According to Sinclair, Jones, and Daley's (2004), there are two types of collocations such as significant collocations and casual collocations. Adopting their significant collocations, one type of significant collocations co-occur with a node within multiple spans. The other type of collocations co-occur with a node within a left and a right position. I call the former window collocations and the latter adjacent collocations.

The notion of collocation is conceptualized by the quote of Firth's (1957) "you shall know a word by the company it keeps." This famous quote has drawn a lot of attentions for the study of collocations. To

10) Francis' approach is a type of corpus-driven approach. The study dealt with what people regularly talk about. Example sentences are as follows: "*I haven't the faintest idea who they are; I can't for the life of me understand what it is you see in it; What's that doing in there? Get it out?; But I can't help feeling that there must be more to it than that; I got feeling something was going on.*" Francis intended to discover speech acts which are everyday sociolinguistic acts.

provide a more practical point of view, Sinclair et al. (2004) discuss a methodology for statistical calculation of significant collocations and collocational patterns of selected lexical items in a certain window span. This study defined collocation as the co-occurrence of two words within a specified environment. They defined three important concepts such as a node, a span, and a collocate. A node is a word which co-occurs with other words. A span is a distance between a node and its co-occurred words. A collocate is a word which co-occurs with the node within the a specified distance.

2.4.2 Statistical Measures for Window Collocations

Evert (2004) extensively explored background theories and tested various statistical measures for finding collocations. His study suggested that old and new hypotheses which each of the statistical measures assumes have to be tested under a range of conditions. More recently, Gries (2013) pointed out that there are three criteria for finding collocations: (1) directionality, (2) dispersion, (3) type-token distribution among collocates. Directionality means that the degree of strength between collocates. For example, a second word, *course,* in *of course* shows a more intense relationship with *of*, but *of* does not a good cue for *course*. Dispersion indicates the degree of distribution among given texts in corpora. For example, the term, *of course,* appeared in a number of different texts. Third, type-token distribution, it measures that how many different types can be collocates which compete with *course* for a position around *of*.

Jhang and Lee (2014) tested several statistical hypothesis tests with a one million word corpus of maritime English to find out which test shows a better result. They experimented proper cut-off span positions using left four to right four spans or a left five to right five spans. Their study employed both hypothesis testing such as log-likelihood, z-test, and t-test and information theory testing such as mutual information (MI) and mutual information 3 (MI^3) to investigate what statistical measures are effective for finding ESP collocations. They concluded that z-score was a better measure than the others in their corpus.

z-score (Oakes, 1988) assumes the normal distribution which is also known as the z distribution. It is calculated as follows:

(2) $$z = \frac{x - u}{o}$$

In this equation, x is each value in the data, u means the mean of all values in the data and o indicates the standard deviation of the population. I will use this z-score in finding collocates because it will be shown in Chapter 4 that z-score is the most proper statistical measure in this thesis.

2.4.3 Problems of Previous Collocation Analysis

Due to some important variables such as window spans, frequency, and statistical measures, there are no tools which can evaluate all of the

measures that have been developed by Natural Language Processing (NLP) researchers (Church and Hanks, 1990; Dunning, 1993; Michelbacher, 2011; Gries, 2013). It seems that finding collocations depends on the researchers' knowledge and a series of decisions by considering variables such as window spans, frequency, statistical measures, etc.

Therefore, it is important for individual researchers to test all of the statistical measures to find out the most proper methods in general corpora as well as specialized corpora. In this thesis, I will test what kinds of collocation measures might be the best for my ESP corpus.

2.5 Visualization in Corpus Linguistics

In corpus linguistics, a visualization method is a useful way of providing important insights into semantic relationships.[11] In this section, I introduce previous studies of visualization methodologies. I present several text visualization studies which focus on the frequency of words. And then I

11) On the other hands, in the field of cognitive linguistics, a tree diagram has been used for a way of representing some hierarchical nature of any structures. Tree structure visualization has been a very useful method to represent a nature in many disciplines: Biology as in evolution trees, Business in pyramid selling schemes, Computer Science as in binary search trees, Linguistics as in phrase structure trees, and Mathematics as in Von Neumann universe. In modern linguistics theories, tree diagrams are concerned with the representation of sentence structures in Transformational Generative Grammar (TG), Generalized Phrase Structure Grammar (GPSG), Head-Driven Phrase Structure Grammar (HPSG), Role and Reference Grammar (RRG), and so on. Another visualization for linguistic theories was to represent in Arc Pair Grammar which derived from a syntactic theory of Relational Grammar (Perlmutter, 1980). This Arc Pair grammar has been concerned with grammatical relations. The other case is Isa diagram which has been used to represent the theory of Word Grammar (Hudson, 2007).

introduce collocation network studies which focus on semantic relationships between nodes and their collocates.

2.5.1 Text Visualizations

Collocate Clouds (Beavan, 2008) presented the collocates of a word by ordering them alphabetically and altering their font size and brightness, as in Figure 2.1.

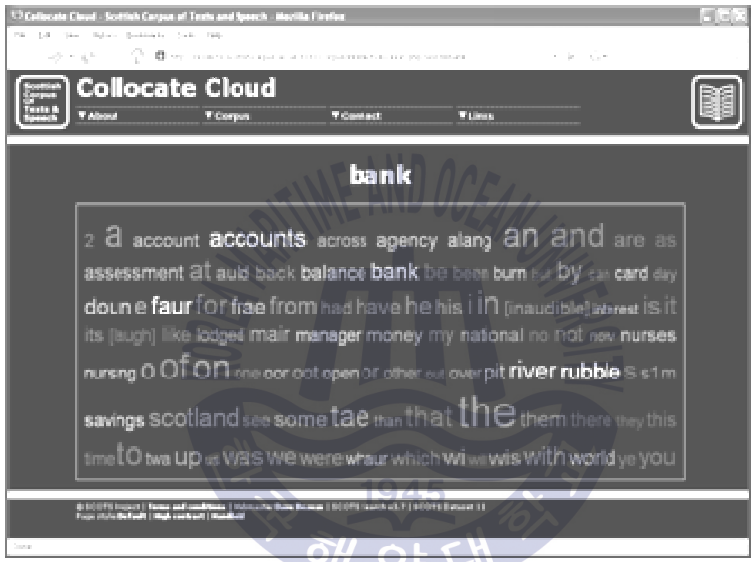


Fig. 2.1 Collocate clouds visualization

Rayson and Mariani (2009) created keyword clouds in which higher keyness words were printed larger so that researchers could visually identify the gradual change of trend for corpus studies, as in Figure 2.2.

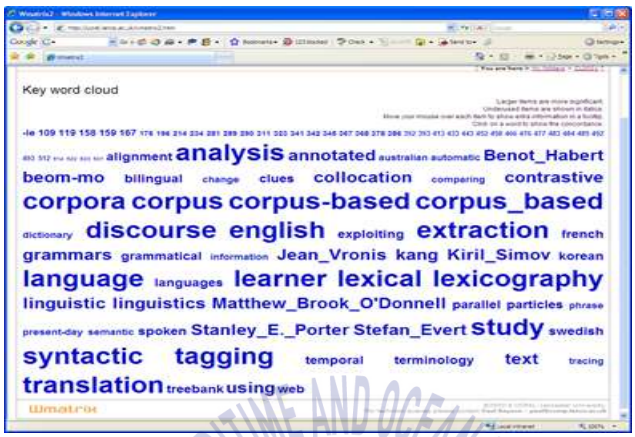


Fig. 2.2 Keyword clouds appeared in 2001 proceedings

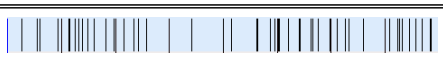
WordSmith 6.0 (Scott, 2015) also creates visualization similar to the previous studies, by automatically highlighting repeated patterns in the surrounding context in (a) and shows word clouds in which frequent words are printed larger and take a center position in (b), as in Figure 2.3.

Fig. 2.3 Concordance and frequent word cloud visualization

Scott (2010) used plot information in his study of Shakespeare's plays. He obtained key clusters from the play "Hamlet" and compared them with all of the other clusters of Shakespeare's plays using WordSmith tools. He studied the place of their locations and found that a nunnery showed a tendency to be located in Act Ⅲ and that "Good My Lord" is rather equally distributed. Thornbury (2010) also suggested that corpus tools could visually display a position of words to predict where certain expressions might appear. Using the plot option in WordSmith tools, Thornbury presented frequent four-word clusters and selected *embarrassed* from a four-word cluster, *I was so embarrassed*, and showed that *embarrassed* occurs at the last part of the text.

In ESP studies, Jhang and Lee (2013b) visualized clusters by using WordSmith Tools 6.0, as shown in Table 2.2.

Table 2.2 Comparison of dispersion plots between two corpora

Function	Examples	Plot (overall)	
Stance bundles	*it should be noted*	MECO II	
		BNC Baby	
Discourse organizers	*shall be provided with*	MECO II	
		BNC Baby	no occurrence
Referential expressions	*in accordance with the*	MECO II	
		BNC Baby	
Special bundles	*international convention for the*	MECO II	
		BNC Baby	no occurrence

2.5.2 Collocation Networks

Williams (1998) explored collocation networks to visualize a Web of interlocking patterns of biology terms in a corpus of plant biology articles. In addition, Williams (2002) visualized DNA with the immediate statistical collocates which reveal the thematic environment of DNA as a technical word with a 560,000 word biology corpus. He explained this type of visualization in accordance with ESP vocabulary learning, as seen in Fig 2.4.

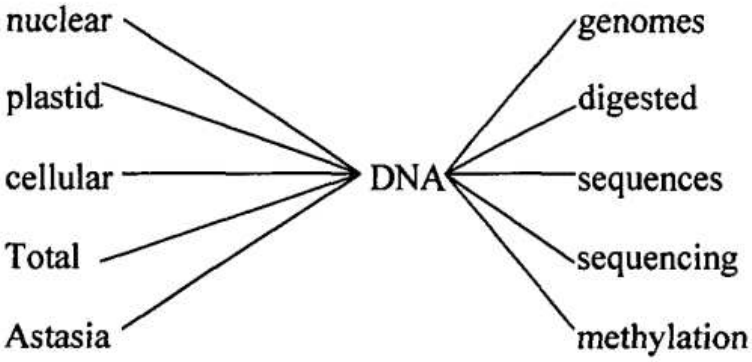


Fig. 2.4 Immediate statistical collocates of DNA

McEnery (2006) employed a visualization technique that draws collocation networks, as in Figure 2.5. This shows keywords linked by the common collocates.

Fig. 2.5 Collocation networks

Stuart and Botella (2009) studied knowledge networks of specific science and technology discourse communities with regard to keywords and clusters as regards their distributions across text plots and discipline levels, as in Figure 2.6.

Word	Abstract	Introduction	Methods	Results	Conclusion
25.found	94	82	133	116	181
26.solution	148	301	273	206	226
27.observed	90	123	**191**	**214**	178
28.interaction	75	27	45	59	77
29.polymeric	**35**	15	12	14	37
30.immunosensors	22	21	10	9	24
31.immunosensor	25	10	14	15	28
32.solvents	33	29	34	23	29
33.concentration	78	108	**118**	**150**	128
34.range	79	96	**100**	**118**	79
35.samples	104	184	168	132	210
36.liquid	66	71	39	42	43
37.solutions	82	153	85	74	52
38.mobility	39	11	19	16	14
39.adsorbed	22	26	31	22	25
40.reported	75	66	68	75	78
41.prepared	60	111	37	19	32
42.pesticide	25	14	12	10	11
43.immunoassays	**29**	12	10	6	10
44.measured	63	114	**120**	89	52
45.immunoassay	25	4	9	17	18
46.buffer	17	58	73	46	23
47.binding	**62**	27	18	46	26
48.pesticides	**42**	9	7	5	11
49.concentrations	22	47	**74**	**77**	32
50.calibration	14	27	**31**	**26**	26

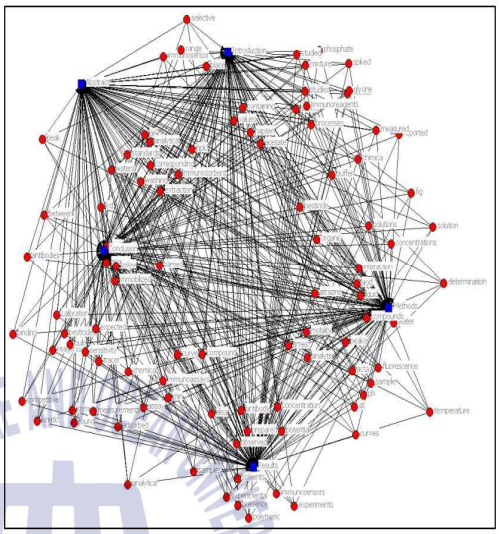


Fig. 2.6 Network examples of keywords per document section

Alonso, Millon, and Williams (2011) used "The DicSci Organic E-Advanced Learner's Dictionary of Verbs in Science" for a vocabulary visualization. They adopted a corpus-driven approach to a new dictionary by using collocation networks, as seen in Figure 2.7(b).[12)]

12) He advocated some advantages of visualization on collocational studies, claiming "[T]his widened network thus gives the semantic environment of the keyword within a biomedical context and also helps to differentiate between potential synonyms."

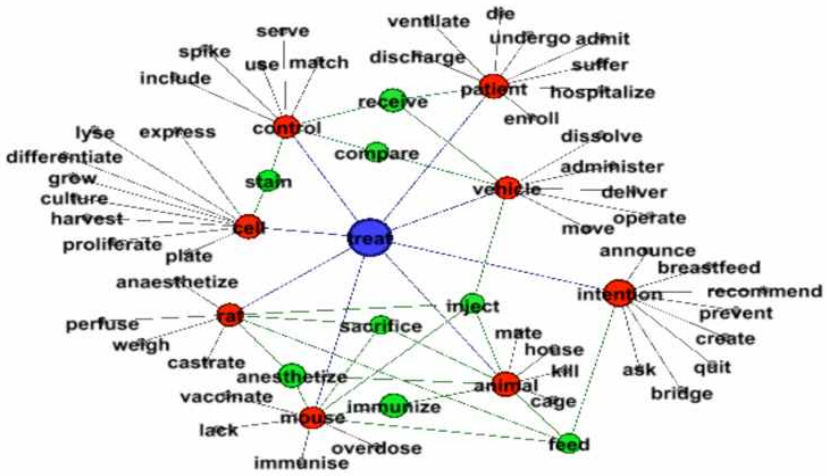


Fig. 2.7 Collocational network from the verb *to treat*

Moretti (2013) adopted a computational approach to deal with huge literary corpora by using z-score and social network analysis to investigate Shakespeare's plays and other novels. In "Distant Reading", he used network visualization among characters and argued for his finding from the character networks of "King Lear", as seen in Figure 2.8.

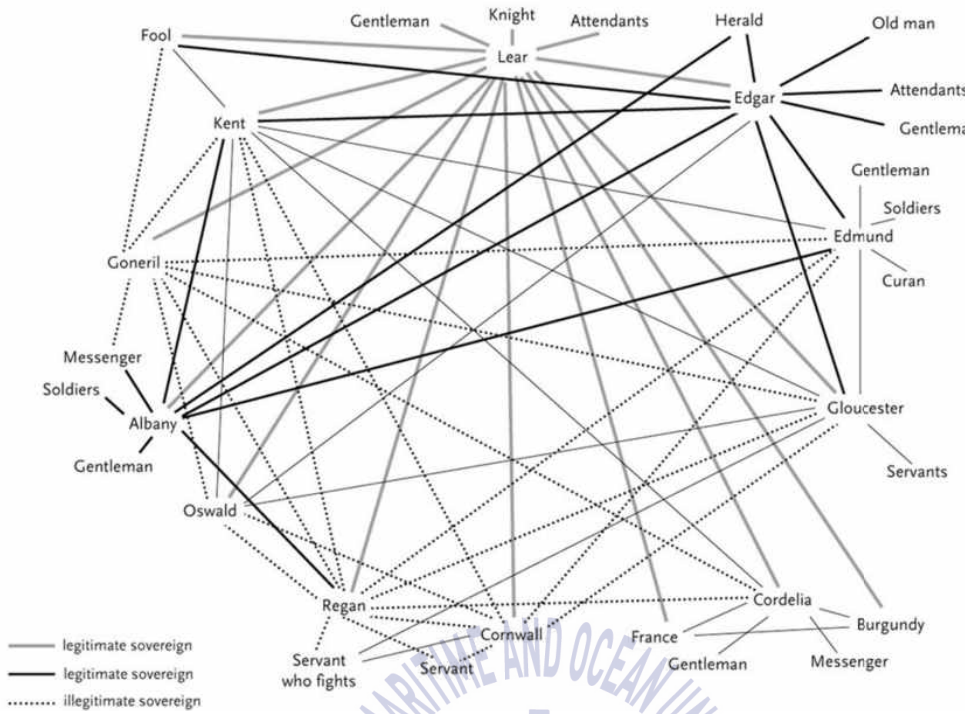


Fig. 2.8 Character Network of "King Lear"

Recently, Jhang and Lee (2013a) visualized collocation networks of near-synonyms of *maritime-marine* and *ship-vessel* in the MEC and explained their semantic relations. They focused on the strength of MI^3 scores by presenting diverse thickness and marking numbers along with node words, as seen in Figure 2.9.

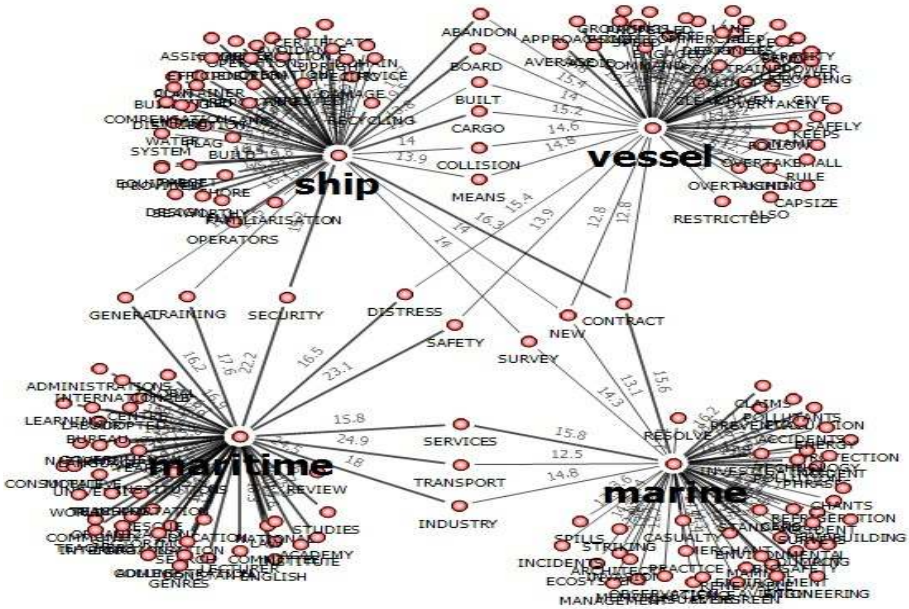


Fig. 2.9 Network of near-synonyms of *maritime-marine* and *ship-vessel*

As a semantic network, Faber, León-Araúz, and Reimerink (2014) drew conceptual networks of *groyne* which shows semantic relations among words, as seen in Figure 2.10.

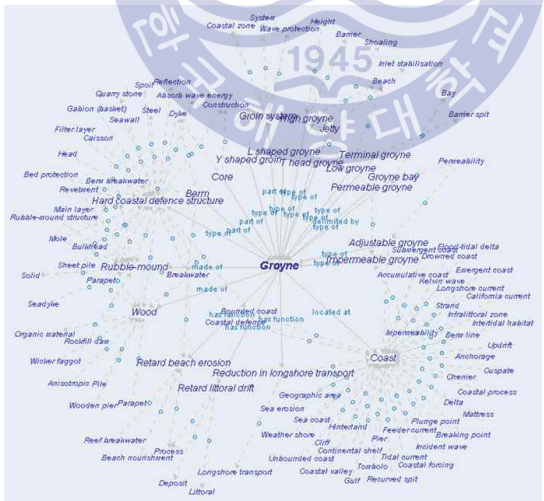


Fig. 2.10 Conceptual map of *groyne*

I think that the most important contribution to visualization of corpus linguistics seems to be a study conducted by Brezina, McEnery, and Wattam (2015) collocation networks including various options for operationalization of the network construction by developing GraphColl 1.0[13] software which makes a systematic analysis of collocation networks available. They revisited collocation networks based on McEnery's (2006) study of the Society for the Reformation of Manners Corpus (SRMC) by using GraphColl to demonstrate collocation networks by suggesting insights into the relationship between meanings. Their study employs different statistical measures for identifying collocations. It deals with directional collocation networks based on Delta P, as shown in Figure 2.11 (Brezina et al., 2015: 158).

13) The software provides 12 different statistical measures for identifying collocations and window span options. With regard to collocation network constructions, a node word and its collocates are drawn based on the criteria of collocation measures. In addition, the software enables users to move beyond first-order collocates to find connections. These interesting options of their research imply that it is very important to explore multi-connected patterns of collocation networks for providing a further explanation of meaning relationships.

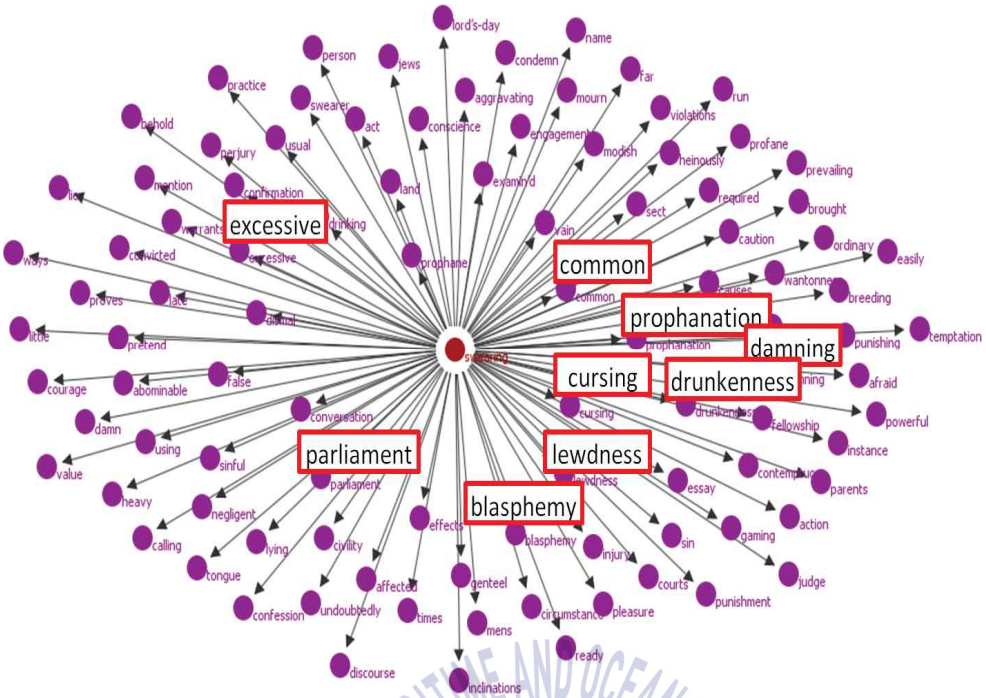


Fig. 2.11 Collocates of *swearing*: replication of McEnery (2006)

More recently, Margan and Meštrović (2015) described LaNCoA, a language network construction and analysis toolkit implemented in Python. It provides co-occurrence networks, syntactic networks, and shuffled networks by using syllable networks and grapheme networks. It also provides several text manipulation methods such as corpora cleaning, lemmatization, and stop words removal.

2.6 Language Networks

In this subsection, I review basic concepts of keywords linkages for keyword networks for the debate on language networks, comparing keyword

networks with collocation networks which have been preferred to prominent network research with a collocation point of view mentioned in the previous subsection. Then I examine important previous studies and definitions of concepts necessary to network structures for language networks.

2.6.1 Basic Concepts

With regard to language network constructions, Scott and Tribble (2006: 75) adopted basic concepts of four types of link of keywords such as strings, stars, cliques, and clumps cited from Jones (1971: 56), as shown in Figure 2.12.

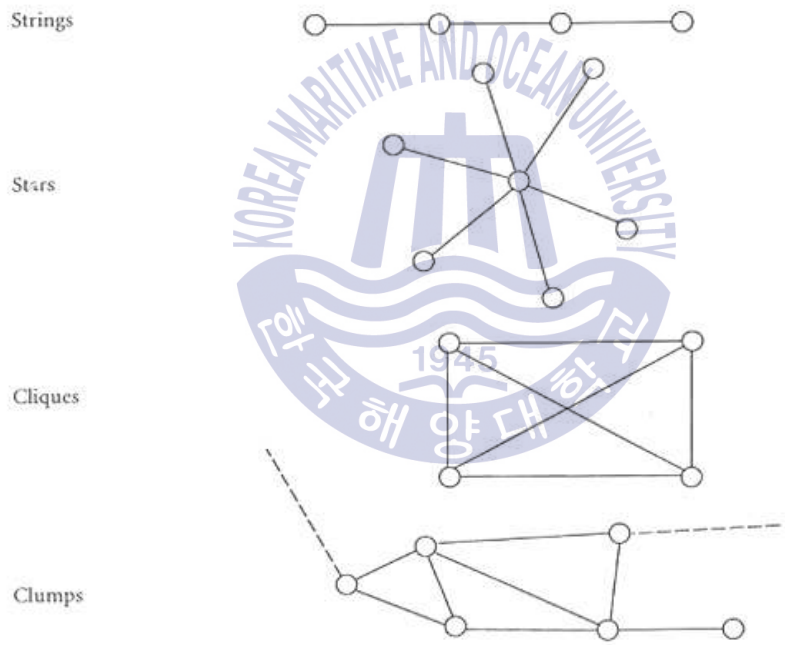


Fig. 2.12 Four types of link of keywords

Expanding Jones' idea, Scott and Tribble (2006: 76) developed a keyword linkage network between keyword nodes and linked keywords connecting her four types of keywords with one combined group, as shown in Figure 2.13.

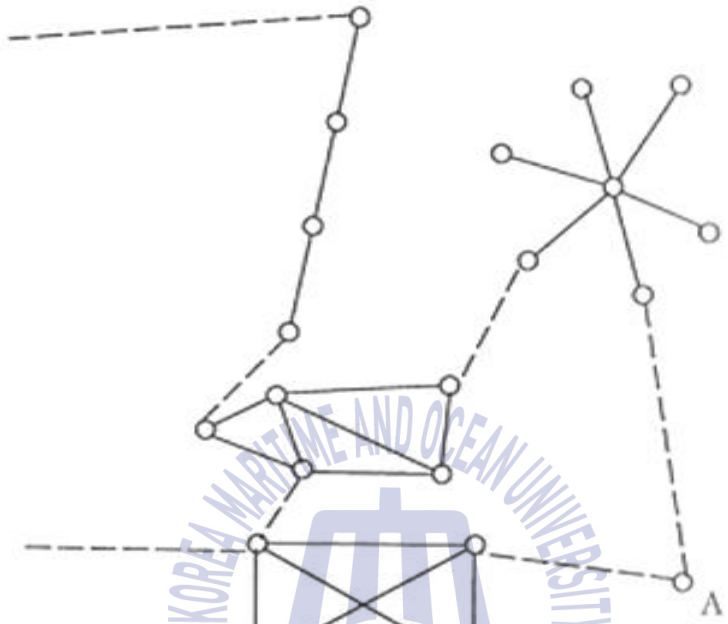


Fig. 2.13 Keyword linkage network

As seen in Figure 2.13 above, linked keywords are defined as a set of keywords which co-occur with a certain keyword within a specified window span. A window span is defined as the number of word-forms before and after a position of the keyword as a node.

As noted in Stubbs (2001: 29), a node is the word-form or lemma which will be investigated. Therefore, I call language networks consisting of keywords and their linked keywords keyword networks in this thesis. On the other hand, collocation network has also studied as a kind of language network analysis, as discussed in the previous subsections. Therefore, I will explore some characteristics of collocation networks by comparing them with keyword networks in Chapter 4.

2.6.2 Previous Studies

My language network analysis complies with the complexity of language systems through network analysis, providing significantly statistical properties which can discover a new type of universals found in human language similar to the universals found in other academic fields such as Physics, Biology, Social Sciences, etc.

Language network analysis uses language texts as the data source and analysis the content of texts (Sole, Corominas, Valverde, and Steels, 2005; Lee, 2014). Language network analysis seems to be a branch of studies for social network analysis.[14] The main goal of social network analysis detects and interprets patterns of social ties between actors (Nooy, Andrej, and Vladimir, 2011). When a language network is built, researchers can analyze

14) NetMiner 6.0 and UCINET 4.0 are widely used for social network analysis. LEXIMAPPE is also used as a co-occurrence analysis tool. LexiURL is specialized as a Website link analysis tool. NetDraw is famous for a visualization tool. Recently, NodeXL is widely used for that purpose. More recently, GraphColl is a free software which is a multi-platform tool for the analysis of linguistic collocation networks developed by Lancaster University in 2015.

the language data using a number of algorithms used for social network analysis. Social network analysis has been adopted to analyze complex relationships to find hidden meaning (Watts and Strogatz, 1998; Scott, 2000; Newman, 2001; Barabasi, 2002; Lee, 2012). Social network studies have investigated how network structures are linked and what these structures really mean. Knowledge network is involved with authors, keywords, and references for network modeling. Social network analysis has studied elements which comprise human societies such as individuals and organizations. In order to model things or abstract phenomena, social network analysis research uses the terms such as network, node, and link in computer science terminology. These terms are also called graph, vertex, and edge in Mathematics.

There are some studies incorporating human language corpora and social network analysis, as discussed in Masucci and Rodgers's (2006) Orwell's 1984 novel, Zhou, Hu, Zhang, and Guan's (2008) Chinese language using "People's Daily" corpus, and Liang, Shi, Tse, Liu, Wang, and Cui's (2009) Chinese and English essays, novels, popular science articles, and news reports. Many of these researches utilized a variety of network measures such as the nearest neighbors, diameter, average degree, degree distribution, clustering coefficients, and average shortest path length, focusing on the characteristics of power law (Newman, 2005) or Zipf's law (Zipf, 1949)[15].

15) According to definitions of power law and Zipf's law retrieved by a Wikipedia Website (https://en.wikipedia.org/wiki), power law means that a change of one quantity causes a proportional change in the other quantity. Similarly, Zipf's law means that a word's frequency is inversely proportional to its rank. Since a corpus data complies with power law or Zipf's law distribution patterns, my data can be a subject of language network analysis.

Language network can be constructed based on the co-occurrence relationship. The collocation used in this language network is one of the most important subjects in corpus linguistics. It is worth mentioning that there are a few studies for a review of domestic language network researches. There are several studies; Jung and Kang (2012) for occupation nouns in the newspaper articles, Lee and Kang (2012) for network analysis of a field of politics, Jhang and Lee (2013a) for visualizing near-synonym networks in maritime English, and Lee (2014) for evaluations of research methods in Korean social network analysis studies.

2.6.3 Definitions

To help understand network analysis, the following definitions of network concepts are necessary to learn, as outlined in Nooy et al. (2011). The following definitions within quotation marks are borrowed in their book "Exploratory Social Network Analysis with Pajek". These definitions will be used in Chapter 4.2 to describe a framework of network analysis.

(3) Definitions of network concepts for network analysis

a. Definition of a graph: A graph is a set of vertices and a set of lines between pairs of vertices.

b. Definition of a simple undirected graph: A simple undirected graph contains neither multiple edges nor loops.

c. Definition of the degree of a node: The degree of a node is the number of lines incident with it.

d. Definition of one-mode network: In a one-mode network, each vertex can be

related to every other vertex.

e. Definition of two-mode network: In a two-mode network, vertices are divided into two sets and vertices can only be related to vertices in the other set.

f. Definition of geodesic: It is the shortest path between two vertices.

g. Definition of betweenness centrality: The betweenness centrality of a vertex is the proportion of all geodesics between pairs of other vertices that include this vertex.

h. Definition of eigenvector centrality: The eigenvector centrality of a vertex is the extent to which it is linked to vertices with high eigenvector centrality.

i. Definition of a strong component: A strong component is a maximal strongly connected subnetwork.

j. Definition of a clique: A clique is a maximal complete subnetwork containing three vertices or more.

2.6.4 Types of Language Network Constructions

The theoretical background for language network was extensively discussed by Sole et al. (2005). They presented three kinds of language networks. The first type of network constructions is to use co-occurrence networks. A co-occurrence network is constructed when two node words co-occur within at least one sentence, as shown in Figure 2.14 (Sole et al., 2005: 2).

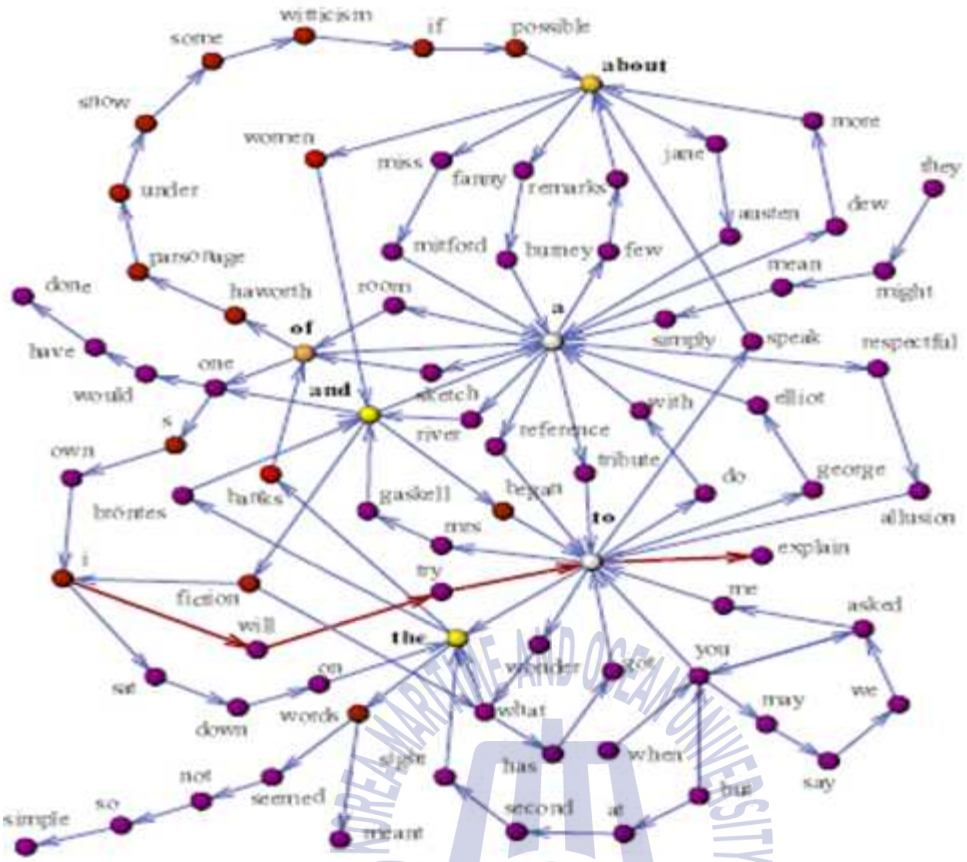


Fig. 2.14 Co-occurrence network construction

The second type of network constructions is to use syntactic networks. This network uses constituent structures based on dependency syntax (Sole et al., 2005: 4). The third type of network constructions is to build up semantic networks. This network is constructed with basic semantic relations: synonymy, hypernymy, hyponymy, meronymy, and antonymy. In this semantic network, one concept word is linked to another concept word based on semantic relations continuously, as seen in WordNet. Among these

three types of language network constructions discussed above, the first type of co-occurrence networks is very close to my language network analysis to follow in Chapter 4.

Chapter 3. Maritime English Corpus

3.1 Introduction

This chapter discusses the specialized corpus compilation and its analysis of the corpus. It discusses representativeness, balance, and size of the corpus. I describe how to compile the MEC including stratified random sampling, Web crawling and cleaning, and converting PDF to texts. I describe how to build a list of reference multi-word compounds by evaluating two dictionary sources with one online supplementary source in order to reveal how important it is to consider multi-word compounds in ESP corpus studies. And then I present the procedures to tag multi-word compounds using customized Python coding. I also compare maritime English corpus with and without compounds based on the comparisons of basic statistics, word lists, n-gram lists, keyword lists, dispersion plots, and visualization to show that there are significant differences.

3.2 Corpus Design

The MEC is comprised of academy, news, laws and textbooks genres. For selecting raw corpus data, I consulted with many experts employed in maritime communities and maritime-related industry fields through email surveys, face to face interviews, and attendance at maritime workshops.

These experts who have given me many pieces of advice have been working at several different institutions such as the International Maritime Organization (IMO), World Maritime University (WMU), Korea Maritime and Ocean University (KMOU)[16], Mokpo National Maritime University, Korean Register of Shipping (KR), Korea Maritime Institute (KMI), and Korea Institute of Maritime and Fisheries Technology.

I decided to compile a four million word corpus, selecting equal amounts of words from four different genres which represent academy, news, laws, and textbooks. There are some practical considerations to determine the size of the four million word corpus. I considered running time for corpus data using personal computers in the data-driven learning (DDL) environment. When teaching students focusing on DDL methods, the corpus size influences teaching and learning. If students are engaged in extracting keywords, key word linked list, and n-grams, it takes much time to get the results if the corpus contains more than four million words. In addition, a sub-corpora of the MEC can be compared with a sub-corpora of BNC Baby in order to find the characteristics of an ESP genre because BNC Baby consists of academic writing, newspaper texts, imaginative writing, and spontaneous conversation and has almost equal amount of size for each genre which has one million words, representing the full BNC. Moreover, around four million word corpus is appropriate for language network analysis because this size of corpus produces a proper number of keywords, linked keywords, and collocates to enable software program to visualize and

16) There is a previous study for a compilation of a small size corpus of maritime English (Hong and Jhang, 2010).

compute network analyses algorithms. Based on these reasons, the MEC size is decided to be comprised of four million words and each of the four genres consists of a one million word sub-corpus.

To collect data for an academic genre, I used Springer's database (http://www.springer.com) which provides numerous journals to scientific and professional communities and Elsevier's Science Direct (http://www.sciencedirect.com) which is one of the largest publishers in the world. I selected the most relevant maritime related academic journals such as "Maritime Policy and Management", "Journal for Maritime Research", "Maritime Studies", "Gyroscopy and Navigation", "Aegean Review of the Law of the Sea and Maritime Law", and "WMU Journal of Marine Affairs". All articles in these journal lists were saved manually as PDF files, as shown in Table 3.1.

Table 3.1 List of academic journal sources

Text_ID	Titles	Sources
A01	Maritime Policy and Management	http://www.tandfonline.com/toc/TMPM20/current#.Vb8g7O2wfIU
A02	Journal for Maritime Research	http://www.tandfonline.com/toc/rmar20/current#.Vb8hFe2wfIU
A03	Maritime Studies	http://www.maritimestudiesjournal.com/
A04	Gyroscopy and Navigation	http://www.springer.com/engineering/mechanical+engineering/journal/13140
A05	Aegean Review of the Law of the Sea and Maritime Law	http://www.springer.com/law/international/journal/12180
A06	WMU Journal of Marine Affairs	http://www.wmu.se/publications/wmu-journal

A news genre consists of official institution texts and commercial news texts. Official institution sources are “IMO Press Briefings” and “World Maritime University News”. Commercial news contains specialized maritime news Websites which are regarded as hub sites by the experts. These sources are “World Maritime News”, “The Maritime Executive”, “Marinelink”, and “Maritime Today News.” Since these Websites contain numerous articles, we used a Wget crawler to collect them. After the collection, an NLP Python program automatically extracted only sentences out of these texts. Table 3.2 shows a list of news Web-site sources.

Table 3.2 List of news website sources

Text_ID	Websites	Sources
N01	IMO Press Briefings	http://www.imo.org/MediaCentre/PressBriefings
N02	World Maritime News	http://worldmaritimenews.com/archives
N03	Marinelink	http://www.marinelink.com/
N04	World Maritime University	http://www.wmu.se/news
N05	Maritime Today News	http://www.maritimetoday.com/
N06	The Maritime Executive	http://www.maritime-executive.com/offshore-news

A law genre is a collection of the IMO regulations and codes recently released by the IMO. In order to collect database of formal regulations and codes, I obtained an agreement with KR which allows me to use the IMO official legal texts for academic purposes. Thus, the IMO data could be inserted to the law genre with its permission. KR's department which is in charge of the IMO official legal texts provided some of these data as a form of CD UNIX forma. Table 3.3 shows a list of maritime law sources.

Table 3.3 List of maritime law sources

Text_ID	Titles	Sources
L01	AFS 2001	http://www.krs.co.kr
L02	Bunker 2001	http://www.krs.co.kr
L03	BWM Convention	http://www.krs.co.kr
L04	COLREG 2014 Consolidated Edition	http://www.krs.co.kr
L05	FSS Code 2014 Consolidated Edition	http://www.krs.co.kr
L06	FTP Code 2014 Consolidated Edition	http://www.krs.co.kr
L07	IBC 2014	http://www.krs.co.kr
L08	IGC 2014	http://www.krs.co.kr
L09	III Code	http://www.krs.co.kr
L10	IMDG Code 2014	http://www.krs.co.kr
L11	ISM Code	http://www.krs.co.kr
L12	ISMBC Code 2014 Consolidated Edition	http://www.krs.co.kr
L13	ISPS Code	http://www.krs.co.kr
L14	LSA Code	http://www.krs.co.kr
L15	MARPOL 2014 Consolidated Edition	http://www.krs.co.kr
L16	MLC 2014 Consolidated Edition	http://www.krs.co.kr
L17	Noise Code	http://www.krs.co.kr
L18	RO Code 2014 Consolidated Edition	http://www.krs.co.kr
L19	Ship Recycling 2014 Consolidated Edition	http://www.krs.co.kr
L20	SOLAS 2014 Consolidated Edition	http://www.krs.co.kr
L21	STCW Convention & Codes	http://www.krs.co.kr
L22	TONNAGE 1969	http://www.krs.co.kr

Maritime-related textbooks are selected for the last genre. I considered to include various fields so the contents of selected books are economics,

safety, transport, history, policy, etc. The number of collected textbooks is 30 kinds and all of them were PDF formats. Later, these PDF files are transformed into txt files and then they are filtered and extracted by an NLP process. Table 3.4 shows a list of book sources.

Table 3.4 List of textbook sources

Text_ID	Titles	Sources
T01	A Global Union for Global Workers: Collective Bargaining and Regulatory Politics in Maritime Shipping	Routledge
T02	Admiral Lord Keith and the Naval War Against Napoleon	University Press of Florida
T03	Maritime Communities and Vegetation of Open Habitats	Cambridge University Press
T04	Command of the Sea	Charles Scribner's Sons
T05	International Maritime Transport	Routledge
T06	Island Disputes and Maritime Regime	Springer
T07	Jurisdiction and Arbitration Clause	Springer Berlin Heidelberg
T08	Maritime Delimitation	Martinus Nijhoff
T09	Maritime Economics	Routledge
T10	Maritime Fiction Sailors and the Sea	Palgrave Macmillan
T11	Maritime Law and Policy in China	Routledge-Cavendish
T12	Maritime Safety Law	Springer
T13	Maritime Security in the South China Sea	Ashgate
T14	Maritime Security	Routledge
T15	Maritime Transportation Safety	Routledge
T16	Maritime Work Law Fundamentals: Responsible Ship owners, Reliable Seafarers	Springer
T17	Oceans Governance	Allen & Unwin
T18	Places of Refuge for Ship	Martinus Nijhoff Publishers Boston

T19	Random Seas and Design of Maritime	World Scientific Publishing Company
T20	Review of Maritime Transport 2006	United Nations
T21	Roots of Strategy Book 4	Stackpole Books
T22	Security for Airport and Aerospace, Maritime and Port, and High-Threat Targets in Belgium	ICON Group International
T23	State Responsibility for Interferences with the Freedom of Navigation in Public International Law	Springer
T24	Sustainable Maritime transportation and Exploitation of Sea Resources Proceedings of the 14th International Congress	CRC Press
T25	The Carriage of Dangerous Goods by Sea	Springer Berlin Heidelberg
T26	The Evolving Maritime Balance of Power in the Asia-Pacific	World Scientific Pub Co. Inc.
T27	The Maritime Dimension of International Security	RAND Corporation
T28	The Maritime Engineering Reference book	Butterworth-Heinemann
T29	The Unforgiving Coast Maritime	Oregon State University Press
T30	Towards Principled Oceans Governance	Routledge

The sum of these collected data is much more than 400 million words. The following chapters describe how to collect texts and how to compile each sub-corpora by using NLP.

3.3 Corpus Compilation

The section illustrates how to collect texts from many different sources and how to extract only texts from these sources using Python and other software tools employed.

3.3.1 Stratified Random Sampling

The random sampling method is necessary to ensure both representativeness and balance of a corpus. For representativeness of a corpus, it is important that the data should be selected evenly from a lot of collected texts. For the balance of a corpus, it is important to collect the exact number of words in accordance with the stratified sampling method. In this vein, to create a four million word corpus, I have to sample each genre consisting of one million words. Since there are no commercial tools which will make available the above purposes, a customized Python program was used for this process. The first thing to extract sentences using this program is to set up directory paths for input text files and output text files. The second step is to decide how many words are extracted by inserting the number of words into WANT_WORDS_CNT, as shown in Figure 3.1.

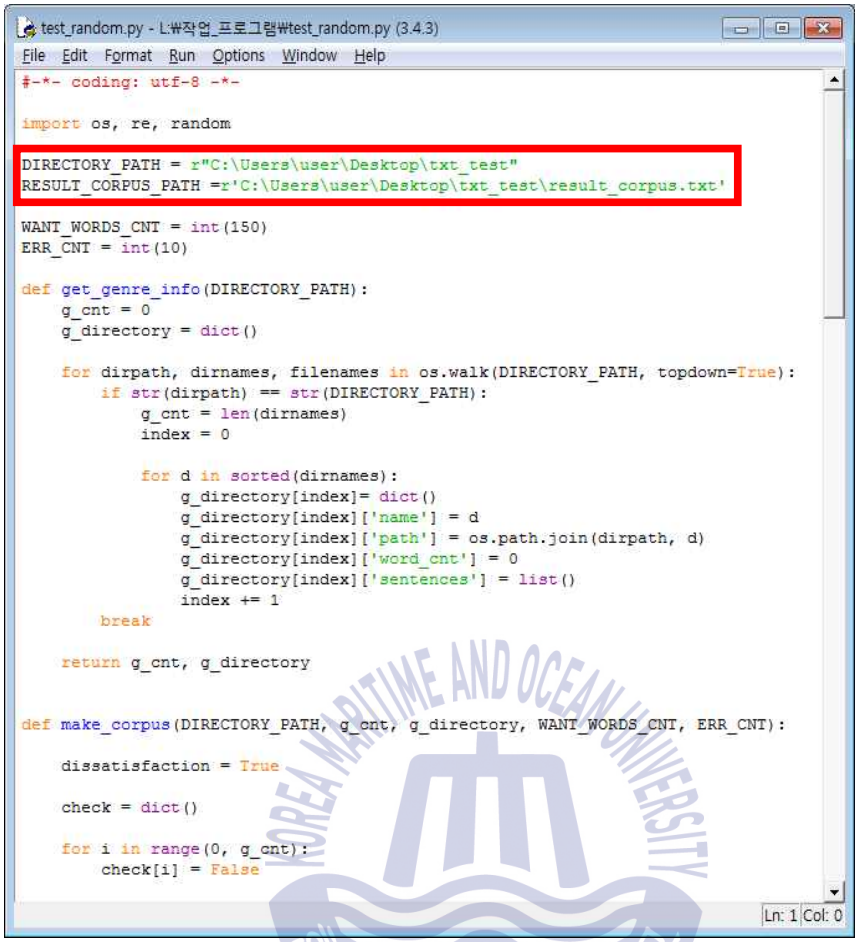


Fig. 3.1 Python coding for random sampling

3.3.2 Web Crawling and Cleansing

Numerous news articles are collected from many Websites. One of good Web crawler software programs, which is called Wget (cf. http://www.gnu.org/software/wget/), was used to collect online news articles. The crawling steps are as follows. First, a DOS commander, cmd was typed and then Wget was activated by commanding “wget -r -l 5 -np -A

aspx -T 1500 -t 5 −c + Website sources". Figure 3.2 shows the images of Wget setup and crawling Websites.

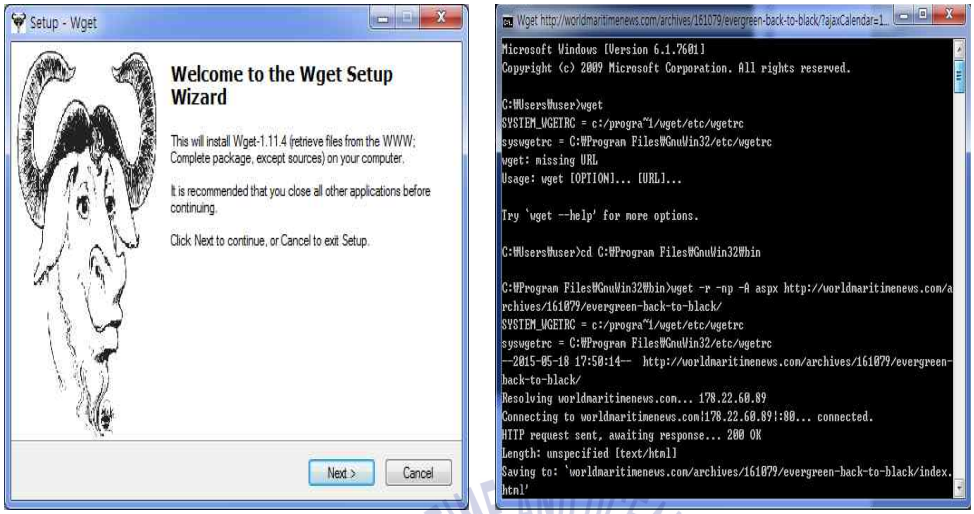


Fig. 3.2 Images of wget setup and crawling websites

The next step is to extract only sentences from the collected HTML documents by using Python program coding. The Python code is designed to implement two purposes. One is to eliminate all the non-alphabetical characters, unnecessary symbols, and images, as seen in Figure 3.3.

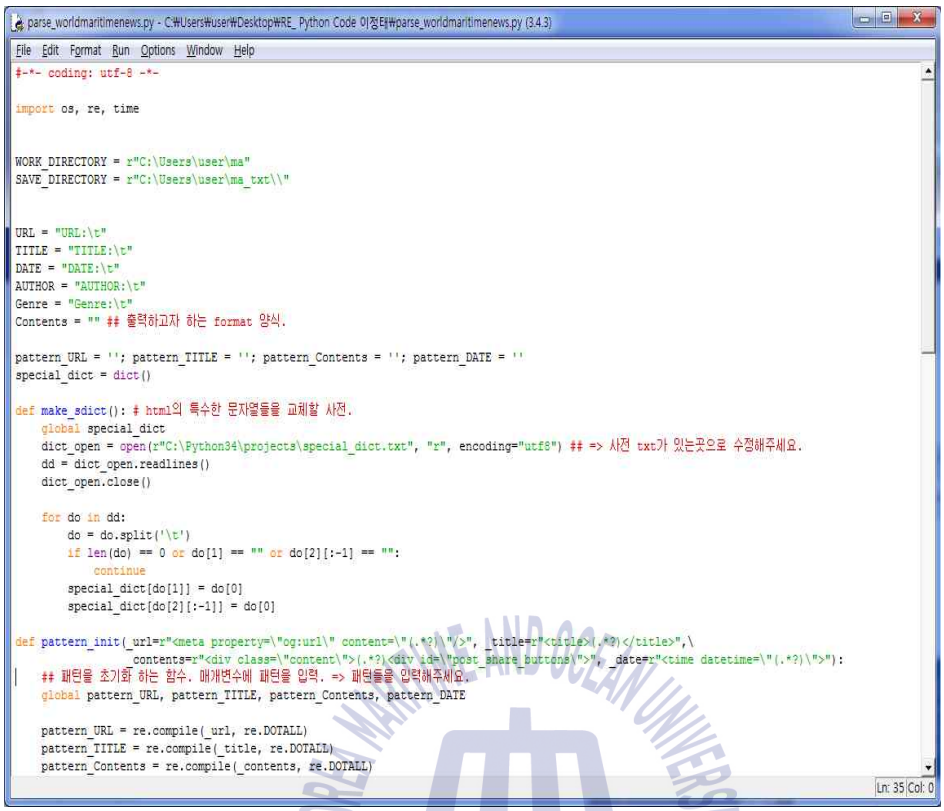


Fig. 3.3 Python news article extraction coding

The other is to collect meta information closely related to the articles. For example, the Python program automatically finds meta information of a URL, a title of an article, a date of written articles, an author or authors, and a genre. Figure 3.4 shows an output of a news text extracted from “Maritime Executive”.

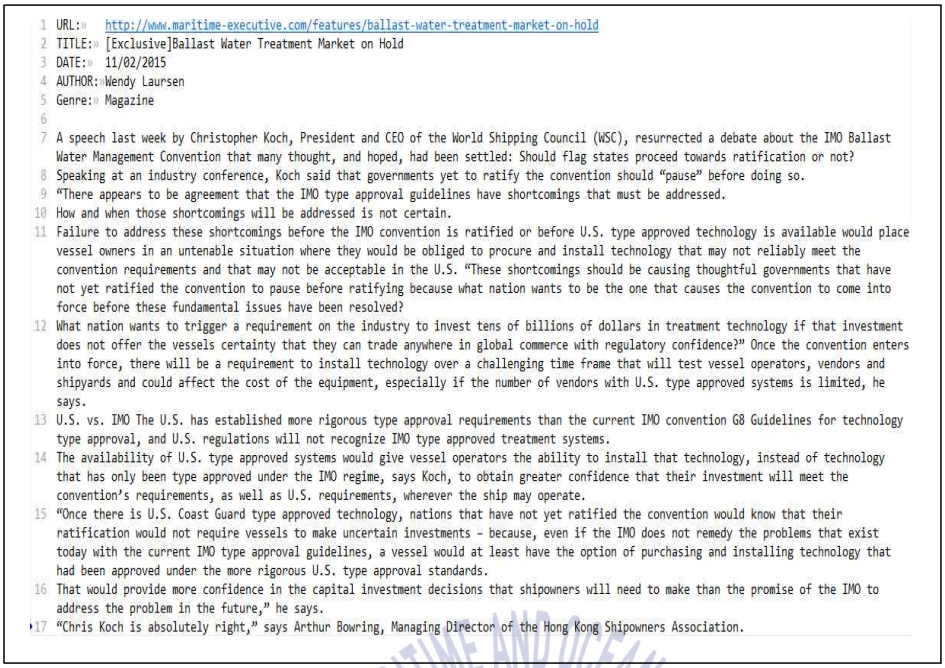


Fig. 3.4 An output of a news text extracted

3.3.3 Converting PDF to Texts

Most academic articles were collected as a Portable Document Format (PDF). After transforming PDF files to text files, we extracted sentences from these texts. In this stage, a free software program called nPDF (cf. http://iblogbox.com/npdf/) was used to transform PDFs to texts because the nPDF program provides options for the user to select only texts. Although the nPDF is used for collecting texts, there are still unwanted remainders such as non-alphabetic symbols and html tags. Thus, I used Python coding to extract sentences. Figure 3.5 shows Python coding, an extracting process and how to save the results in the designated folders.

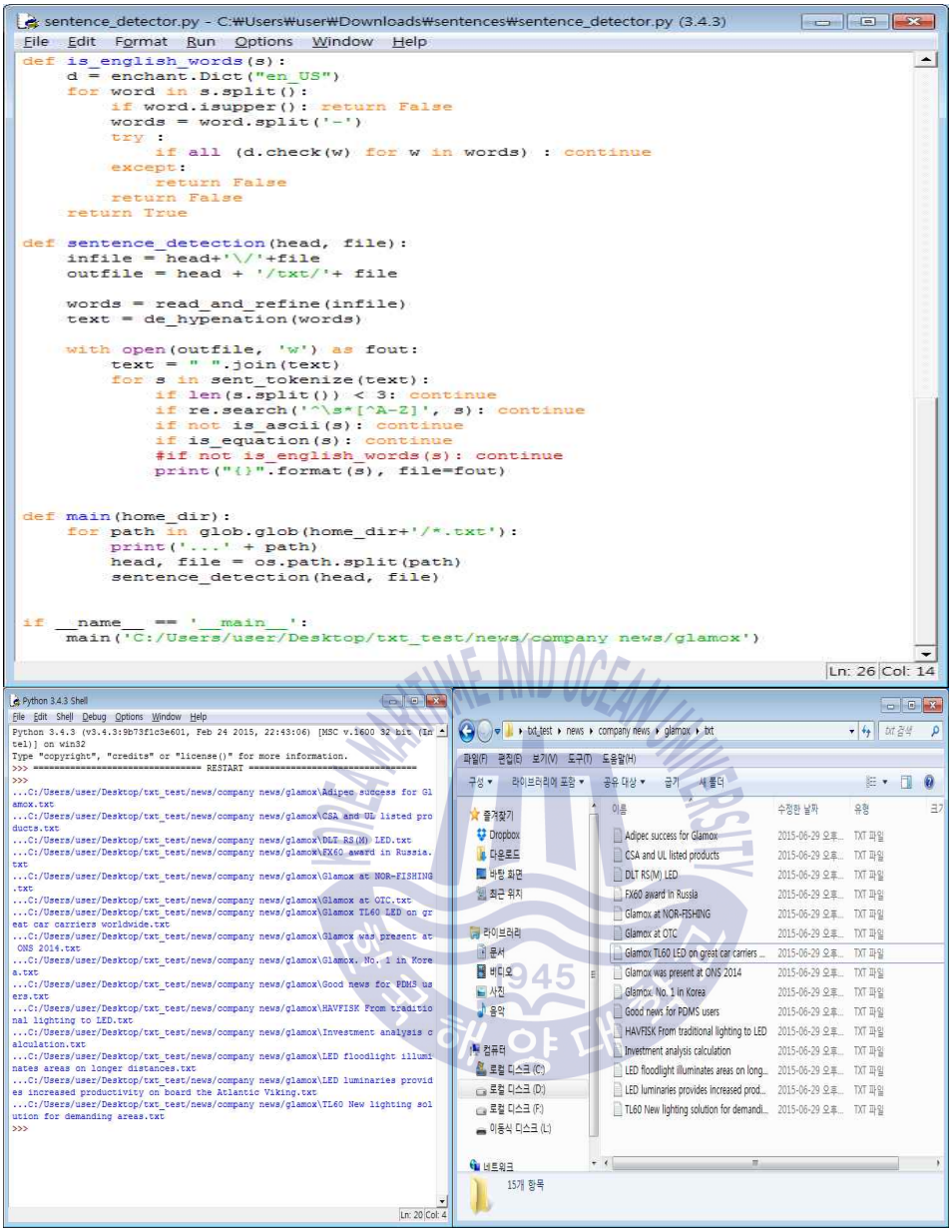


Fig. 3.5 Python coding and the output

3.4 Multi-word Compounds

I struggle to tackle an unsolved issue of corpus linguistics in relation to vocabulary research. Maritime English vocabulary dictionaries include a large number of compounds such as *air draft, ballast water, bow thruster, chief officer, starboard bow*, etc. However, there are no software tools (e.g., even in most notable tools such as AntConc, WordSmith Tools, and Wmatrix) for finding these compounds automatically. Therefore, I take this issue in my maritime English research and develop an NLP platform in which multi-word compounds can be automatically found.

One of the most important purposes of this research is to include multi-word compounds in the corpus. Finding maritime English compounds is very important when extracting a list of words. Current corpus tools such as AntConc and, WordSmith tools cannot detect automatically multi-word compounds in case more than two words are separated between spaces. This is a serious problem unsolved in the general and ESP corpus vocabulary researches.

I compared three ways of spelling English compounds (Plag, Braun, Lappe, and Schramm, 2009: 100) with maritime English compounds, as shown in Table 3.5.

Table 3.5 Types of general English compounds and maritime English compounds

	General English	Maritime English
1	*ashtray, windmill, hotline*	*seafarer, shipyard, offshore, shipbuilding*
2	*fast-food, icy-cold, call-girl*	*Hapag-Lloyd, double-hull, ABS-classed*
3	*ice cream, income tax increase*	*coast guard, ballast water management*

One type of compounds consists of single vocabulary, and another type consists of two words linked between dash mark, the other type consists of two or three words which are separated by spaces. My thesis concerns the third type of multi-word compounds such as coast guard, ballast water management, etc.

So far, there are many different approaches to extracting compounds with different languages. To find out recent research trends, I reviewed papers from The 2013 NAACL HLT (Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies), The 2014 EACL (Conference of the European Chapter of the Association for Computational Linguistics) multi-word expression workshops. There are four major areas such as a statistical basis, a linguistic basis, a machine learning basis, and complex approaches. Table 3.6 summarizes different approaches.

Table 3.6 Different approaches to compound extraction

Simple Statistics Approaches	● Smadja (1993): Chi-square test, t-test, z-test, log-likelihood ratio ● Kunchukuttan and Damani (2008): Point-wise MI, log-likelihood, frequency of occurrence ● Zhang, Yoshida, Tang, and Ho (2009): Enhanced MI based association measures ● Tanmoy (2013): Extraction of multi-word expressions using Chunk information and various algorithms.
Linguistics Approaches	● Rayson, Archer, Piao, and McEnery (2004): The UCREL semantic analysis system ● Tanmoy, Das, and Bandyopadhyay (2011): A semantic clustering based approach ● Mahesh and Sinha (2011): Replicating words, pair of words, samaas (N+N, A+N), Sandhi, Vaalaa morpheme
Machine Learning Approaches	● Wanger (2000): Enriching a lexical semantic net with selectional preferences ● Pecina (2008): Linear logistic regression, linear discriminant analysis (LDA) Neural Networks
Complex Approaches	● Sag, Baldwin, Bond, Copestake and Flickinger (2002): Maintaining the right balance between symbolic and statistical approaches ● Dandapat, Mitra, and Sarkar (2006): Statistical Investigation of Bengali Noun-Verb (N-V) Collocations as Multi-word expressions ● Ramisch, Caseli, Villavicencio, Machado, and Finatto (2010): Combines the strengths of different sources of information using a machine learning

Although these previous studies provide a number of algorithms to find compounds, there are no perfect solutions for identifying multi-word compounds in ESP vocabulary studies. However, I found one useful and practical method which uses reference dictionaries for identifying multi-word compounds. Thus, I selected two maritime English dictionaries and one supplementary source to tag multi-word compounds within the corpus: "Dictionary of Maritime and Transportation Terms" (Monroe and Stewart, 2005) and "International Maritime Dictionary" (Kerchove, 1961). In order to compensate the most contemporarily added terms, I used one supplementary source from Wikipedia's "Glossary of Nautical Terms (2015)", https://en.wikipedia.org/wiki/Glossary_of_nautical_terms.

3.5 Critical Evaluation and Tagging for Multi-word Compounds

Using these three sources, I collect all entry items listed in each dictionary and then combine them to make a list of all entry items including single and multi-word compounds to create a reference multi-word compound list. First, in order to create each list of all entry items from two published dictionary books, I type all entry items by using a Microsoft excel program. However, selecting entry items from dictionaries is not just straightforward process due to different writing system for word formation. "Dictionary of Maritime and Transportation Terms" expressed frequently used head-words in a different way compared with actual language use. For example, a multi-word compound *deck* is expressed as a combination of *deck* with a comma and it's followed adjectives such as *deck anchor, deck*

boat, and *deck bulkhead.* In this type of word formation, I changed these multi-word compounds as *anchor deck*, *boat deck,* and *bulkhead deck* in accordance with actual language use in maritime domains. Second, to build a word list of Wikipedia's "Glossary of Nautical Terms", I copied the data from the Web and then pasted them by using a free software, notepad ++ for further cleaning processing. To accelerate the cleansing process, I erased unnecessary lines and symbols by using regular expressions. I obtained a list of all entry items from the source.

The basic statistics of each list of all entry items from these three sources is as follows. First, "Dictionary of Maritime and Transportation Terms" has 6,265 terms, "International Maritime Dictionary" has 11,044 terms, and "Glossary of Nautical Terms" has 1,297 terms. The total number of token of entry items is 18,636. After getting rid of overlapped entry items, the total number of type of entry items is 16,729. A list of all entry items combined with three sources contains various multi-word compounds ranging from bi-grams to nine grams, as shown in Table 3.7.

Table 3.7 Percentage of multi-word compounds in a reference multi-word compound list

N-grams	Types	%	Examples of Multi-word Compounds
Single words	5,697	34.05	*bow, starboard*
2-grams	8,770	52.42	*harbor master*
3-grams	1,654	9.89	*bow chock plate*
4-grams	475	2.84	*anchor by the stern*
5-grams	86	0.51	*dry bulk self unloader ship*
6-grams	31	0.19	*global maritime distress and safety system*
7-grams	11	0.07	*left-hand draft in this set of marks*
8-grams	3	0.02	*between the devil and the deep blue sea*
9-grams	2	0.01	*International Convention for the Safety of Life at Sea*
Total	16,729	100	

As seen in Table 3.7, it is striking that there are only 8,770 2-gram compounds (52.42%). These 2-grams indicate that the corpus vocabulary word list should consider multi-word compounds when extracting word lists. Especially, ESP vocabulary list should include multi-word compounds in order to correctly count the proportions of specific purpose vocabulary items. The second dominant type is 5,697 single vocabulary items (34.05%). The third dominant type is 3-grams (9.89%) comprising 1,654. Compared to the number of 2-grams, the number of 3-grams is quite reduced. Other types are 475 4-grams (2.84%), 86 5-grams (0.51%), 31 6-grams (0.19%),

11 7-grams (0.07%), 3 8-grams (0.02%), 2 9-grams (0.01%). This finding is also quite impressing because 11,032 entry items (65.95%) are multi-word compounds. Therefore, the statistics shows how much multi-word compound tagging is important in ESP studies.

Table 3.8 displays the top 20 terms from a reference multi-word compound list showing the top and last terms out of all entry items.

Table 3.8 Top 20 and last terms out of all entries on a reference multi-word compound list

No	Dictionary of Maritime and Transportation Terms	No	International Maritime Dictionary	No	Wikipedia's Glossary of Nautical Terms
1	a1	1	aak	1	abaft
2	aa	2	aalboot	2	abaft the beam
3	ab	3	aback	3	abandon ship
4	abaft the beam	4	abaft	4	abeam
5	abandon	5	abaft the beam	5	abel brown
6	abandoned goods	6	abandonment	6	able seaman
7	abandonment	7	abandonment clause	7	aboard
8	abatement	8	abeam	8	above board
9	abc analysis	9	able seaman	9	above-water hull
10	abel tester	10	aboard	10	absentee pennant
11	ablation	11	above deck girder	11	absolute bearing
12	able-bodied seaman	12	abox	12	accommodation ladder
13	able-bodled seaman	13	abreast	13	accommodation ship
14	about ship	14	absence flag	14	accommodation hulk
15	aboveboard	15	absolute contraband	15	act of pardon
16	abovedeck	16	aburton	16	act of grace
17	abrasion	17	acceleration	17	action stations
18	abreast	18	accident boat	18	admiral
19	absent flag	19	accommodation	19	admiralty
20	absolute accuracy	20	accommodation berth	20	admiralty law
.	.	.	.	.	.
6,265	zone price	11,044	zee bar	1,297	yawl boat

The above table is worth consideration for compound tagging. A large portion of dictionary entries consists of multi-word expressions. Maritime dictionary entries include a bunch of multi-words separated by spaces, as exemplified in the three sources as *able-bodied seaman, abaft the beam, abandon ship* respectively. Therefore, it is not possible for us to extract these multi-word compound entries from any other corpora, because all of the available corpus analysis tools do not provide a list of compound words. They can help extract only single word or words between dash marks.

In order to prepare multi-word compound tagging, I combined all the entry items from three different sources to make a reference maritime English dictionary. Then I divided it into two word lists. One is a list of single words which I call a reference single word list. The other is a list of multi-word compounds which I call a reference multi-word compound list. All these three kinds of word lists are shown in Table 3.9.

Table 3.9 All the entry items listed from three different sources

No	All the Entry Items	No	Single Words	No	Multi-word Compounds
1	a1	1	a1	1	abaft the beam
2	aa	2	aa	2	abandon ship
3	aak	3	aak	3	abandoned goods
4	aalboot	4	aalboot	4	abandonment
5	ab	5	ab	5	abandonment clause
6	aback	6	aback	6	abc analysis
7	abaft	7	abaft	7	abel brown
8	abaft the beam	8	abandon	8	abel tester
9	abandon	9	abandonment	9	able seaman
10	abandon ship	10	abatement	10	able-bodied seaman
11	abandoned goods	11	abeam	11	about ship
12	abandonment	12	ablation	12	above deck girder
13	abandonment clause	13	able seaman	13	above-water hull
14	abatement	14	aboard	14	absence flag
15	abc analysis	15	abovedeck	15	absent flag
16	abeam	16	abox	16	absentee pennant
17	abel brown	17	abrasion	17	absolute accuracy
18	abel tester	18	abreast	18	absolute bearing
19	ablation	19	absorption	19	absolute contraband
20	able seaman	19	aburton	19	absolute pressure
.	.	.	.	.	.
16,730	zulu	8,583	zulu	8,147	zone price

For the extraction method, I programmed Python codes in order to create a multi-word compound tagged corpus. First, I tokenized the study corpus to make a word list, lemmatized the word list using TreeTageer, and then found all kinds of types of lemmatized words. Second, I lemmatized the reference multi-word compound list. Third, I matched two lemmatized word lists. Fourth, I found all types of multi-word compounds. Finally, I combined a new reference multi-word compound list including lemmatized words and all types of words from lemmatized words. This new reference multi-word compound list was used to tag the MEC. The whole process described above is seen in Figure 3.6.

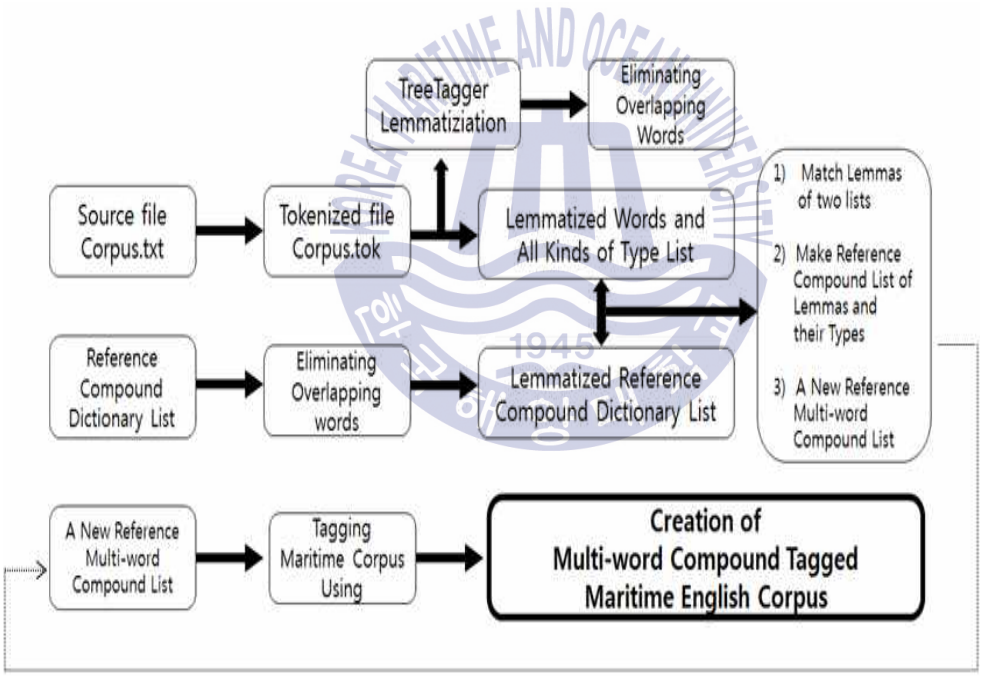


Fig. 3.6 Creation of multi-word compound tagged MEC

en_route means that the ship is underway at_sea on a course or courses, including deviation from the shortest direct_route, which, as far as practicable for navigation purposes, will cause any discharge to be spread over as great an area of the sea as is reasonable and practicable.
except as provided in paragraph 2 of this regulation, in ships delivered after 31 December 1979, as defined in regulation 1.28.2, of 4,000 gross_tonnage and above other than oil_tankers, and in oil_tankers delivered after 31 December 1979, as defined in regulation 1.28.2, of 150 gross_tonnage and above, no ballast_water shall be carried in any oil fuel tank.
where the need to carry large quantities of oil fuel render it necessary to carry ballast_water which is not a clean_ballast in any oil fuel tank, such ballast_water shall be discharged to reception_facilities or into the sea in compliance with regulation 15 of this annex using the equipment specified in regulation 14.2 of this annex, and an entry shall be made in the oil_record_book to this effect.

Fig. 3.7 Result of compound tagged specific vocabulary terms

Multi-word compounds are expressed by adding '_' symbols. For example, *significant_wave_height* is tagged through two '_' symbols. There are words as *en_route, at_sea, direct_route, oil_tankers, gross_tonnage, ballast_water, clean_ballast,* and *oil_record_book.* The below excerpt from part of maritime law texts shows how this tagging system works, as shown in Figure 3.7.

3.6 Comparison of With and Without Compounds

Comparisons of with and without compounds are discussed in the following sections in terms of basic statistics, word lists, n-gram lists, and keyword lists.

3.6.1 Comparison of Basic Statistics

This section compares word lists with and without compounds. Table 3.10 shows the statistics before tagging multi-word compounds and Table 3.11 indicates the statistics after tagging multi-word compounds. It is necessary to mention the different kinds of definitions for words for understanding basic statistics. Token means the total number of words. Type is the number of distinct words in a corpus. Lemma is an abstract form. It stands for the class of words with stem. Finally, word family means the base form of a word plus its inflected forms and derived forms made from affixes, such as inflections including third person *-s,* past *-ed,* progressive *-ing*, plural *-s* possessive *-s*, comparative *-er* and superlative *-est*, and derivations including *-able, -er, -ish, -less, -ly, -ness, -th, -y, non-, un-, -al, -ation, -ess, -ful, -ism, -ist, -ity, -ize, -ment, in-,* etc. In this study, token and type are mostly used in the statistics.etc. In this study, token and type are mostly used in the statistics.

Table 3.10 Statistical results before tagging multi-word compounds

Genres	Token	Type	Type/Token Ratio (TTR)	Standardised TTR
Academy	989,943	79,639	8.04	56.17
News	997,835	39,790	3.98	55.03
Laws	985,621	15,500	1.57	30.05
Textbooks	989,610	53,617	5.41	54.11
Total	3,963,009	138,389	3.49	48.87

Table 3.11 Statistical results after tagging multi-word compounds

Genres	Token	Type	Type/Token Ratio (TTR)	Standardised TTR
Academy	985,284	80,317	8.15	56.31
News	983,350	42,792	4.35	55.74
Laws	971,471	16,433	1.69	30.53
Textbooks	984,180	54,416	5.52	54.34
Total	3,924,285	142,041	3.61	49.29

The observation of these two tables shows that there are common differences between the two results. First, the number of tokens of each genre is reduced when the corpora contain multi-word compounds. Second, the number of types of each of genre is increased after compound tagging. As seen in Table 3.10 and 3.11, the difference of type is due to the newly added multi-word compounds. These findings show that there is significant misunderstanding of an ESP word list if a researcher neglects considering multi-word compounds.

Let us consider some differences found in TTR and STTR. The way to get the TTR is the number of types over the number of tokens multiplied by 100 (TTR = number of type/number of token x 100). When comparing a different size of corpora, the standardised type/token ratio (STTR) is useful. The ratio is calculated for the first 1,000 running words, then calculated afresh for the next 1,000, and so on to the end of a text or corpus. For example, the STTR of a corpus without compound tagging is 48.87. On the other hand, the STTR of a corpus with compound tagging is

49.29. This indicates that the tagged corpus shows a greater variety of vocabulary than without the untagged corpus. Table 3.12 shows a comparison of basic statistics.

Table 3.12 Comparison of basic statistics

Corpora	Token	Type	Type/ Token Ratio (TTR)	Standardised TTR
Corpus without multi-compounds	3,963,009	138,389	3.49	48.87
Corpus with multi-compounds	3,924,285	142,041	3.61	49.29

So far, it has been shown that it is important to consider multi-word compounds in order to provide more accurate word lists in ESP corpus studies. It is important to mention that the basic statistics of a corpus will be significantly different depending on the inclusion of multi-word compounds. From my data, it is found that there are differences in corpus statistics such as token, type, TTR and STTR between two corpora with and without multi-word compounds.

3.6.2 Comparison of Word Lists, N-gram Lists, and Keyword Lists

By comparing four grams between the untagged and tagged MEC word lists, I showed the top 10, middle 5 and last words, as shown in Table 3.13.

Table 3.13 Comparison of the untagged and tagged MEC word lists

N	Untagged Law Corpus	Freq.	N	Tagged Law Corpus	Freq.
1	THE	74169	1	THE	74060
2	OF	46478	2	OF	45993
3	#	35184	3	#	35180
4	AND	28336	4	AND	28232
5	TO	26214	5	TO	26132
6	IN	22031	6	IN	21969
7	BE	17869	7	BE	17869
8	A	17350	8	A	17344
9	OR	16544	9	OR	16497
10	SHALL	16380	10	SHALL	16380
174	WITHOUT	692	174	BALLAST_WATER	686
769	PACKED	165	769	BULK_CARGOES	177
1,777	CARTRIDGE	54	1,777	CHIEF_ENGINEER	54
3,240	SIGNATORY	21	3,240	MOULDED_DEPTH	27
3,863	LEAKAGES	15	3,863	ABANDON_SHIP	15
.	.	.	.	.	.
.	.	.	.	.	.
15,381	YEMEN	1	16,433	YERSINIA	1

The total number of a word list is different between the two methods. It is found that the untagged corpus has 15,381 words and the tagged corpus has 16,433 words. The multi-word compound tagged corpus have more words than those of the untagged corpus.

By comparing MEC between the untagged and tagged corpora, I showed the top 10, the middle and last 4-grams, as seen in Table 3.14.

Table 3.14 Comparison of the untagged and tagged MEC 4-grams

N	Untagged Law Corpus	Freq.	N	Tagged Law Corpus	Freq.
1	IN ACCORDANCE WITH THE	805	1	IN ACCORDANCE WITH THE	805
2	WITH THE REQUIREMENTS OF	401	2	FOR THE PURPOSE OF	388
3	FOR THE PURPOSE OF	388	3	THE PROVISIONS OF THIS	374
4	THE PROVISIONS OF THIS	374	4	THE DATE ON WHICH	326
5	WITH THE PROVISIONS OF	362	5	CONSTRUCTED ON OR AFTER	316
6	THE DATE ON WHICH	325	6	IN THE CASE OF	307
7	CONSTRUCTED ON OR AFTER	315	7	WITH THE PROVISIONS OF	299
8	IN THE CASE OF	306	8	BY THE COMPETENT AUTHORITY	290
9	BY THE COMPETENT AUTHORITY	290	9	WITH THE REQUIREMENTS OF	269
10	TO THE PROVISIONS OF	286	10	ADOPTED BY THE ORGANIZATION	262
11	ADOPTED BY THE ORGANIZATION	262	.	.	.
.	.	.	136	TO THE PROVISIONS OF	78
271	TAKE INTO ACCOUNT THE	56	271	SHIPS BALLAST_WATER AND SEDIMENTS	55
1,839	A MEMBER OF THE	19	1,839	FIT TO PROCEED_TO_SEA WITHOUT	19
4,577	CAPABLE OF WITHSTANDING A	11	4,577	GROSS_TONNAGE CALCULATED IN ACCORDANCE	11
13,454	MODIFICATIONS OF A MAJOR	6	13,454	OF CHIEF_ENGINEER OFFICERS AND	6
19,881	THE HOLDER HAS ATTENDED	5	19,881	THE PORT_STATE AUTHORITY STOWAGE	5
.	.	.	.	.	.
20,753	ZONE IN AN AREA	5	20,643	YOU MAY RE-USE THIS	5

There is a quantitative difference in total numbers. The 4-grams are found in the untagged law corpus as in 20,753 and in the tagged law corpus as in 20,643. On the 4-grams, it is also found that the tagged law corpus contains multi-word compounds such as *ships ballast_water and sediments, fit to proceed_to_sea without, and gross_tonnage calculated in accordance*.

In addition, I compared keywords with and without multi-word compounds, as shown in Table 3.15.

Table 3.15 Comparison of vocabulary on the keyword list

N	Untagged Law Corpus	Keyness	N	Tagged Law Corpus	Keyness
1	SHALL	35,331.09	1	SHALL	35,662.63
2	OR	15,360.23	2	OR	15,518.80
3	BE	10,312.16	3	BE	10,497.82
4	PARAGRAPH	4,423.05	4	PARAGRAPH	4,464.96
5	CONVENTION	4,388.94	5	CONVENTION	4,200.24
6	PROVISIONS	4,106.82	6	PROVISIONS	4,152.81
7	ANY	3,796.81	7	ANY	3,843.97
8	ACCORDANCE	3,795.84	8	ACCORDANCE	3,840.35
9	SPACES	3,729.05	9	PROVIDED	3,278.91
10	CERTIFICATE	3,504.25	10	CERTIFICATE	3,227.68
11	PROVIDED	3,237.54	11	SPACES	3,018.15
33	TANKS	1,986.45	33	DANGEROUS_GOODS	1,998.54
67	CONSTRUCTED	1,139.95	67	BALLAST_WATER	1,116.47
423	UNLOADING	218.61	423	BULKHEAD_DECK	221.39
.	.	.	.	.	.
1,833	EXCEEDED	23.95	1,976	SEDIMENT	23.97
1,834	STORM	-24.08	1,977	BRING	-24.19
.	.	.	.	.	.
2,615	WAS	-3,470.34	2,761	WAS	-3,440.39

To conduct keywords analysis, target and reference corpora have to be defined and prepared. I used a one million word law corpus as a study corpus, and a 3 million word corpus consisting of academy, news, and textbooks as a reference corpus following a cross-validation method. The

statistical criteria used for keyword extraction are that 23.95 or higher critical value and significant level at p<0.000001. As shown in Table 3.15 above, within the top 11 keywords, there is almost no difference on the keyword list. However, because of a lot of multi-word compound keywords on the keyword list of the tagged corpus, the total number of keywords between the untagged and tagged corpora is not the same because there are 2,615 consisting of 1,834 positive keywords and 781 negative keywords in the untagged corpus and 2,761 consisting of 1,976 positive keywords and 785 negative keywords in the tagged corpus.

3.6.3 Comparison of Visualizations

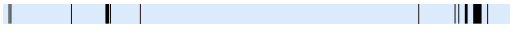
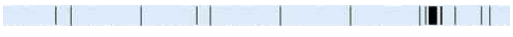
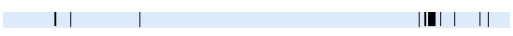
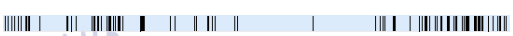
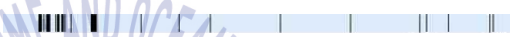
This section deals with two visualization tools such as dispersion plots in WordSmith Tools 6.0 and GraphColl 1.0 to show that tagged multi-word compounds work very well.

3.6.3.1 Dispersion Plots

I attempt to graphically show dispersion plots of keywords in order to show how plot analysis can show the order of themes within a text.[17] Keyword plots are presented in Table 3.16.

17) Plot distribution analysis can be done by using the only text (Scott, 2015)

Table 3.16 Dispersion plots with and without multi-word compound tagging

Words	Corpus	Hits	Per 1,000	Dispersion	Plots
fire	untagged	1207	1.24	0.36	
	tagged	799	0.83	0.39	
lifeboat	untagged	243	0.25	0.21	
	tagged	236	0.25	0.21	
subdivision	untagged	85	0.09	-0.01	
	tagged	81	0.08	-0.01	
stability	untagged	270	0.28	0.78	
	tagged	232	0.24	0.81	
crew	untagged	281	0.29	0.66	
	tagged	268	0.28	0.65	
ballast_water	untagged	749	0.77	0.21	
	tagged	686	0.72	0.21	
bulk_carriers	untagged	81	0.08	0.27	
	tagged	79	0.08	0.28	

Interestingly, these keywords preferred a certain part of the whole text such as beginning, middle, and end of the text. As can be seen in Table 3.16 above, the hits column shows that the keywords without multi-word compound tagging have more occurrences than those with multi-word compound tagging. In addition, almost all of the keywords from untagged and tagged corpora show different dispersion plots.

3.6.3.2 GraphColl 1.0

Multi-word compound words make the GraphColl 1,0 more useful because multi-word compounds such as *port state* and *ballast water* can be visualized in the Graph window. Due to the compound tagger (e.g., under a bar or "_"), these words are expressed as *ballast_water* and *liquefied_gas_tanker*. The underbar tagger enables the software to recognize these multi-words as one single word. As can be seen below, I contrasted two-gram and tri-gram compounds with underbar and without underbar.

In Figure 3.8, there are two types of words such as *ballast water* and *ballast_water* using GraphColl 1.0. (Brezina et al. 2015).

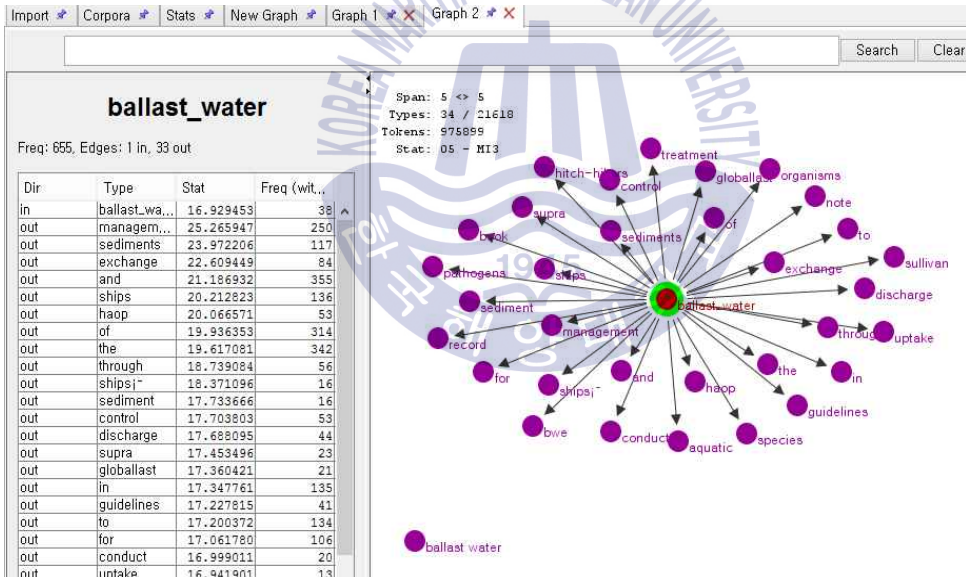


Fig. 3.8 Compound tagging of *ballast_water*

As seen in Figure 3.8, *ballast_water* with compound tagging can show its collocates in visualization, whereas the untagged compound *ballast water* cannot.

In Figure 3.9, I draw two nodes, *ballast_water* and *abandon_ship* together using GraphColl 1.0.

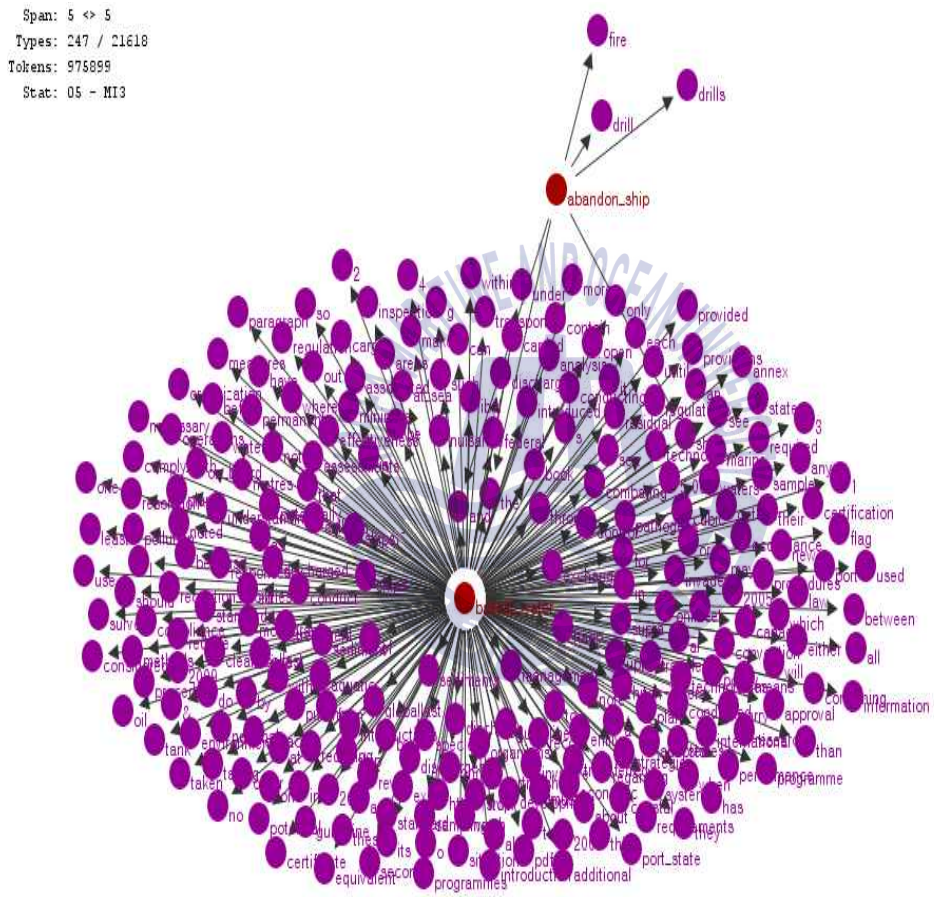
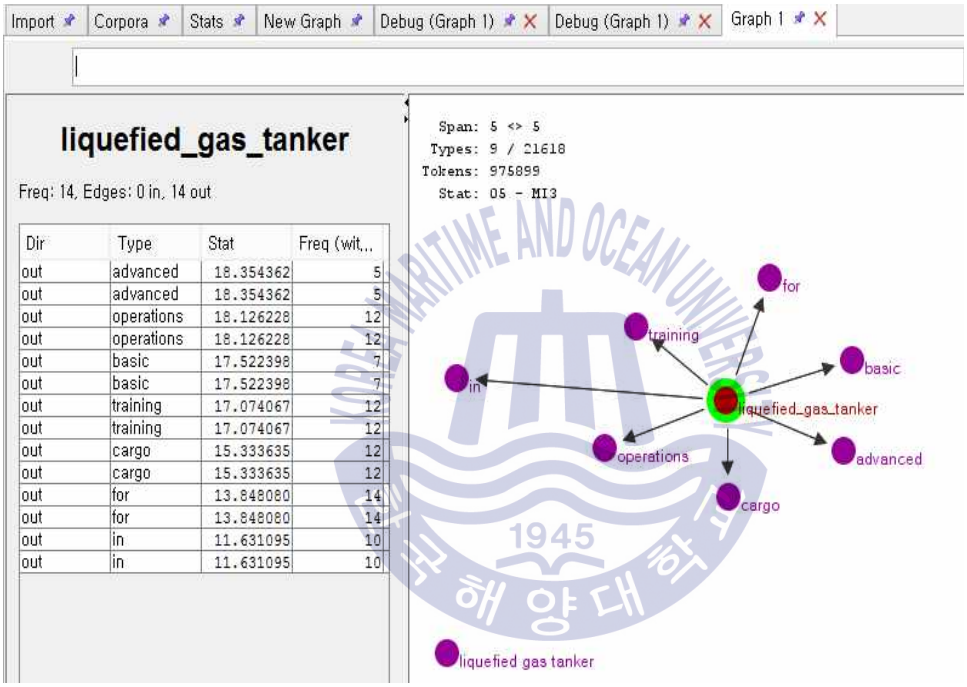


Fig. 3.9 Compound tagging of *ballast_water* and *abandon_ship*

As seen in Figure 3.9, two nodes, *ballast_water* and *abandon_shop* are linked with their shared collocates in the way of collocation networks.

In visualization drawn by GraphColl 1.0, a tagged tri-gram compound example of *liquefied_gas_tanker* can be a node linked with collocates, whereas an untagged tri-gram *liquefied gas tanker* cannot, as seen in Figure 3.10 below.



Dir	Type	Stat	Freq (wit...
out	advanced	18.354362	5
out	advanced	18.354362	5
out	operations	18.126228	12
out	operations	18.126228	12
out	basic	17.522398	7
out	basic	17.522398	7
out	training	17.074067	12
out	training	17.074067	12
out	cargo	15.333635	12
out	cargo	15.333635	12
out	for	13.848080	14
out	for	13.848080	14
out	in	11.631095	10
out	in	11.631095	10

Fig. 3.10 Compound tagging of *liquefied_gas_tanker*

3.7 Summary and Implications

The first part of this chapter described the procedures of compiling the MEC with multi-word compound tagged. I created a four million word MEC consisting of academy, news, laws and textbooks genres. Each genre contains one million words and a maritime law sub-corpus is a study corpus in this thesis. A reference corpus was built with academy, news, and textbooks sub-corpora consisting of 3 million words following a cross-validation method. I proposed a method of tagging English multi-word compounds following two procedures. (1) I make an evaluation of multi-word compounds in my study corpus with a list of multi-word compound entry items collected from the two maritime dictionaries and one supplement Web source. (2) The customized Python programs were used to tag multi-word compounds in my study corpus using underbar.

The rest of this chapter deals with the comparison of the MEC with and without multi-word compounds with regard to STTR, word lists, n-gram lists, keyword lists, dispersion plots, and GraphColl 1.0. The STTR without multi-word compound tagging is 48.87 and that with compound tagging is 49.29. The latter shows a greater variety of vocabulary than the former. The word list without compound tagging has 15,381 words and that of the tagged corpus has 16,433 words. The latter shows more words than the former. The 4-grams of the untagged corpus show 20,753 and those of the tagged corpus have 20,643. Interestingly, the former shows more 4-grams than the latter. In the comparison of keywords with and without multi-word compounds, the untagged corpus has 2,615 keywords and the tagged corpus

has 2,761 keywords. The latter shows more keywords than the former. Almost all of the keywords from untagged and tagged corpora show different dispersion plots. Using the GraphColl 1.0, I showed that tagged multi-word compounds can successfully link their collocates, whereas untagged n-grams cannot. My tagging system works very well in different software.

There are several implications of corpus linguistics point of view. My tagging system for a multi-word compound can help traditional concordance tools and new corpus visualization tools to be developed to a new horizon beyond the current lexico-grammar research. It is because that a corpus tagged with multi-word compounds should reflect authentic language usage for lexical analysis, keyword analysis, and phraseology including collocation and n-grams.

Chapter 4. Language Network Analysis

4.1 Introduction

This chapter proposes a new approach for the incorporation of several methods in corpus linguistics into language network analysis in order to deal with both textual features which keyword networks can show and language features which collocation networks can show. Language network analysis is adopted to provide an explanatory power for corpus linguistic descriptions given in the previous Chapter 3. To do this, I visualize keyword networks which are built with keywords and their linked keywords. The original ideas of keyword networks have been traced to Jones (1971) and Scott and Tribble (2006). I discuss network structures such as centrality and cohesion with regard to eigenvector and betweenness. On the other hand, based on the previous collocation networks which are built through linkages of keywords and their co-occurred collocates (Williams, 1998; McEnery, 2006; Brezina et al., 2015), I hypothesize that collocation networks can be applied to visualizing a specialized corpus in the same way as keyword networks. Hence, I compare keyword networks with collocation networks focusing on network structures such as centrality and cohesion with regard to eigenvector and betweenness in order to explore some differences and similarities between these two network analyses. Finally, I summarize and discuss

pedagogical implications of language network analysis.

4.2 Frameworks of Network Analysis

I discuss a framework of network analysis including choosing source and target nodes, creating two-mode and one-mode data, describing network analysis algorithms, and visualizing network structures.

4.2.1 Source Nodes and Target Nodes

According to Scott (2010), textual features are considered as text-dependent but language features are considered as language-dependent. Hence, I hypothesize that keyword networks and collocation networks are available to explaining descriptions of a whole text or a set of texts for the better understanding of maritime English. My first hypothesis is that keyword networks may show textual features in an ESP maritime corpus. My second hypothesis is that collocation networks may show linguistic features in an ESP maritime corpus. In order to create both keyword networks and collocation networks, it is necessary to determine source nodes and target nodes for the two networks. In this study, a source node refers to a keyword in both keyword networks and collocation networks. A target node refers to a set of linked keywords in keyword networks, whereas it refers to a set of collocates in collocation networks.

To prepare network input data, I need to decide a study corpus and a reference corpus from the MEC for extracting keywords. Based on the

cross-validation method, I choose a one million word maritime law genre which is a sub-corpus of the MEC as a study corpus and a three million word reference corpus which is the sum of the three other sub-corpora such as academy, news, and textbooks genres. The main reason is that maritime law is the most important genre in international maritime community because most of the maritime law texts which I collected are from legal documents released by IMO. For this reason, other genre texts such as academy, news, and textbooks are greatly influenced by the maritime law texts. Another reason is that the maritime law texts contain a variety of compounds compared to other sub-genres. For example, when I present the basic statistical information of a word list, a keyword list, and a linked keyword list, Table 4.1 shows basic statistics of single words and various multi-word compounds.

Table 4.1 Percentage of single words and multi-word compounds in a study corpus

N-grams	Type of Words	%	Keywords	%	Linked Keywords	%
Single words	15,379	93.68	2,594	93.98	1,823	91.76
2-grams	873	5.21	144	5.20	142	7.14
3-grams	151	0.93	21	0.73	21	1
4-grams	17	0.1	1	0.06	1	0.05
5-grams	6	0.04	1	0.03	1	0.05
6-grams	3	0.02	0	0	0	0
7-grams	2	0.01	0	0	0	0
8-grams	1	0	0	0	0	0
9-grams	1	0.01	0	0	0	0
Total	16,433	100	2,761	100	1,988	100

As seen in Table 4.1, the number of type of words is counted on a word list. Keywords are counted on a keyword list. The configuration settings for extracting keywords by using WordSmith Tools 6.0 are a minimum of 10 occurrences of each keyword and Dunning's log- likelihood ratio with a p-value at 0.000001. Types of words are most abundant on a word list, but they are reduced on a keyword list, and more reduced on a linked keyword list. However, I found that the linked keyword list still provides not only single words but also various n-gram types ranging from two to five.

Therefore, I choose linked keywords of the maritime law genre as a study corpus.

The sampling of study keywords as source nodes is based on the proportions of sing words and multi-word compounds in all the entry items from a reference multi-word compound list. It is important to include both single words and multi-word compounds for language networks to investigate how multi-word compounds contribute to language network structures. First of all, I decided to choose 40 keywords as source nodes to show language network structure more clearly. If there are too many nodes, it is difficult to recognize visualization images and also takes too much time for calculation. Therefore, I make a list of source keywords with 20 single words and 20 multi-word compounds to give even portions. To choose multi-word compounds fairly, I considered the percentage of each n-gram on a reference multi-word compound list. Using the statistics of Table 3.7, if all n-gram types are 100%, the total 2-grams (79%) occupy 8,770, and the rest of the n-grams (21%) are 2,262. By following this percentage, I selected 15 2-grams (75%), and 5 multi-grams (25%) including three 3-grams, one 4-gram, and one 5-gram. In addition, due to a lot of candidates from single words and 2-grams, I created a Python random sampling code to sample 20 single, 15 2-grams, and five 3-5 grams keywords. The finally selected items as source keywords are shown in Table 4.2.

Table 4.2 Source keywords for network analysis in a study corpus

	Single words		2-grams		3 to 5-grams
1	ACCORDANCE	1	BALLAST_WATER	1	GENERAL_PURPOSE_CONTAINER
2	ADMINISTRATION	2	BULK_CARGOES	2	NOXIOUS_LIQUID_SUBSTANCES
3	AMENDMENT	3	BULKHEAD_DECK	3	OIL_RECORD_BOOK
4	APPROVAL	4	DANGEROUS_GOOD	4	MEDICAL_FIRST_AID_GUIDE
5	COMPETENT	5	EXPIRY_DATE	5	INTERNATIONAL_CONVENTION_ ON _LOAD_LINES
6	CODE	6	FIRE_ALARM		
7	CONVENTION	7	FREIGHT_CONTAINERS		
8	MAY	8	GROSS_TONNAGE		
9	ORGANIZATION	9	LIQUEFIED_GASES		
10	PACKAGINGS	10	MACHINERY_SPACES		
11	PROVIDED	11	OIL_TANKERS		
12	PROVISIONS	12	ON_BOARD		
13	REGULATION	13	RESCUE_BOATS		
14	REQUIRED	14	STEERING_GEAR		
15	REQUIREMENTS	15	SURVIVAL_CRAFT		
16	SHALL				
17	SPACES				
18	SPECIFIED				
19	SUBSTANCES				
20	TANK				

For finding target nodes for collocation networks, it is necessary to find linked keywords as well. A linked keyword is defined that "[L]inks are co-occurrences of key-words within a collocational span" by Scott (2015). I use the keywords tool out of WordSmith tools 6.0 to extract linked keywords by setting a minimum link frequency of 1 and a link span of five words to the left and five words to the right. I save the linked keywords as an excel file in order to use keywords as source nodes and linked keywords as target nodes.

The 40 keywords have a different number of linked keywords which are produced by Wordsmith tools' link option on the keyword list. I show some examples in Table 4.3. There are different word types of keywords and their linked keywords from single words to 5-grams such as *shall, ballast_water, oil_record_book, medical_first_aid_guide,* and *international_convention_on_load_lines*.

Table 4.3 Five keywords and their linked keywords

Single words	shall (254)	internal, no, treated, burning, establish, operable, hatches, effective, guidance, toxicity, segregated, horizontal, propelling, slop_tanks, apparatus, nozzles, cofferdams, pollution, undue, marked, ships, digits, affixed, moulded, capacitors, general, applied, salvor, interpretations, annotation, described, steel, criteria, superstructures, breathing_apparatus, preferably, specimens, sanitary, commensurate, refer, auxiliary_steering_gear, dose, wearing, general_purpose_container, sheets, decontaminated, drop, circulated, vehicles, metals, manufacturer, annual_survey, sprinklers, bearing, respects, dirty_ballast, gastight, deadweight, checking, fissile-excepted, adversely, criticality, unified, boiler, permeability, solvent, allowed, decease, longitude, inflated, verify, hose, given, thermally, circulate, regard, exempted, masters, consisting, fore-side, flammability, calcium, adsorbed, tank_top, purging, circuits, seafarer, davit-launched, approves, cryogenic, requirement, carriage, correspond, undertakes, familiarization, second_engineer, device, packing_instructions, applies, design, copies, leak, inside, relieving, effect, surface, pipes, nitrates, permanent, shell_plating, inert, vent, document, associated, depleting, deformation, container, double_bottom, columns, invites, malfunction, entrances, oily, movement, outside, aggregate, prevention_of_pollution, frangible, IMDG, oily_mixtures, latitude, programmes, damage_stability, inhibitor, forming, sounding_pipes, deflagrating, slop_tank, uptake, liquefied_gas_tanker, addition, capacitor, indicate, passivation, regarding, control_stations, skin, oxide, delegated, portions, application, specific, located, manner, divisions ...
2-grams	ballast_water (195)	ventilation, issued, applicable, materials, than, weather_deck, foam, annex, prescribed, test, bilges, other, after, contents, appropriate, remanufactured, certification, equipment, accidental, type, walls, liquid, person, authorized, applicator, tanks, cargo, freight_container, unit, door, radionuclide, hydrogen, concerned, readily, subdivided, chemical, engaged, laid, packed, subjected, sill, identification, classified, below, attached, radio, flammable_solids, plastic, flexible, leakproof, transverse, cartridges, tubes, vehicle, non-refrigerated, freeboard, text, label, fire_pump, bar, surfaces, two-thirds, edp, permit, non-conformity, arrangements, ducts, supplemented, special, seagoing, examination, refer_to, organization, witness, source, noise, sound, required, standard, displayed, requirements, account, miscellaneous, fly, approval, paragraph, designated, suitably, non-fixed, deck, certified, space, freeboards, embarkation, provided, relief, declaration, provisions, whenever, segregation, flammable, manoeuvre, present, thermal, regulation, any, list, constructed, deformation, inert_gas, notified, convention, refrigerating, gross, chief_engineer, approving, entitled, carry, means, accession, may, metal, shall, shielding, used ...
3-grams	oil_record_book (18)	flammable, shielding , such, label, comply_with, minimum, isolation, transmitting, ducts, removable, lightest, permissible, compatibility, embarkation, after, purposes, ship, shall
4-grams	medical_first_aid_guide (12)	precautions, manhole, substances, msc, sealed, facility, defined, satisfaction, ensure, permissible, be, by
5-grams	international_convention_on_load_lines (11)	forward_perpendicular, accept, way, binary, obtain, idcm, emerged, immediately, iridium, developer, SOLAS

For finding target nodes for collocation networks, it is necessary to find out suitable statistical measures for extracting collocations. In order to identify the most reliable measure, I compare several different statistical measures such as z-score, t-score, dice, log likelihood, MI, and MI^3, all of which are offered by WordSmith Tools 6.0. Table 4.4 shows the top 20 collocates of each of these measures when *maritime* is chosen as a node word within the window span of five to the left and to the right.

Table 4.4. Statistical comparison of collocations of *maritime* of top 20 words

N	Z-score	T-score	Log-like	Dice	MI	MI^3
1	security	the	the	security	the	the
2	safety	of	of	safety	of	of
3	labour	and	and	international	and	and
4	international	in	in	transport	in	security
5	committee	to	security	industry	to	in
6	epitomized	a	safety	policy	a	safety
7	scandinavian	security	to	and	security	international
8	domain	safety	international	in	safety	to
9	terrorism	international	a	committee	international	a
10	delimitation	for	transport	the	for	transport
11	policy	is	for	labour	is	for
12	transport	by	industry	of	by	labour
13	law-making	on	policy	law	on	policy
14	kongsberg	as	labour	management	as	industry
15	act	transport	by	world	transport	committee
16	security	with	committee	a	with	by
17	high-threat	industry	on	china	industry	on
18	holistic	that	as	organization	that	is
19	fiction	has	is	to	has	as
20	industry	policy	law	has	policy	delimitation

As can been seen in Table 4.4, z-score shows the best results within the top 20 words in that it shows more content words compared to other results. Except for z-score, all other measures have more function words rather than content words. Thus, z-score seems the better statistical test for measuring collocates as supported by the previous study of an ESP collocations (Jhang and Lee, 2014). Considering these results, I choose z-score for finding collocates. z-score values can be positive or negative, showing collocates' strength to node words. In this study, both the z-score value over 3 and P-value less than 0.001 are the criteria for finding more significant collocations by reducing collocates. I found out that the number of collocates to each of the keywords is very different. For example, *cargo* has 15,000 significant collocates, whereas *oil_record_book* has 70 collocates. For this reason, I used the top 50 significant collocates as a cut-off point because they are data enough to show a semantic relationship between keywords and their collocates.

4.2.2 Two Mode Structures and One Mode Structures

In order to visualize language networks, I use NetMiner 4.0 (Cyram, 2015). NetMiner 4.0 has been developed by a Korean company and now its fame is renowned in the world due to user-friendly interfaces. NetMiner 4.0 is used to analyze various kinds of data to detect underlying patterns and structures of the networks. Using an excel file of linked keywords obtained from WordSmith tools, the two mode matrix is input into a NetMiner program. In a two-mode network, source nodes are linked with a number of target nodes so that the entire networks are interconnected through certain

degrees of links among target nodes. Table 4.5 shows two mode data for keyword networks. It shows sources referring to keywords and targets referring to a set of linked keywords.

Table 4.5 Two mode input data for keyword networks

Number	Sources	Targets
1	SHALL	approves
2	SHALL	davit-launched
3	SHALL	damage_stability
4	SHALL	circuits
5	SHALL	purging
.	.	.
254	SHALL	cause
1	BALLAST_WATER	fire_pump
2	BALLAST_WATER	ducts
3	BALLAST_WATER	arrangements
4	BALLAST_WATER	melting_point
5	BALLAST_WATER	party
.	.	.
195	BALLAST_WATER	requirements
1	OIL_RECORD_BOOK	flammable
2	OIL_RECORD_BOOK	shielding
3	OIL_RECORD_BOOK	such
4	OIL_RECORD_BOOK	label
5	OIL_RECORD_BOOK	embarkation
.	.	.
18	OIL_RECORD_BOOK	permissible

This two-mode data is visualized through Kamada-Kawai's (1989) spring embedding algorithm which can effectively visualize the network structures effectively. This algorithm is one of the force-directed graph layout algorithms.[18] The algorithm is as follows:

$$(3) \qquad k_{i,j} = \frac{K}{d_{i,j}^2}$$

K is a constant. $d_{i,j}$ is the shortest path distance between i and j. The algorithm in (3) has some advantages to drawing networks clearly even though they take a little time. The keywords and their linked keywords are visualized in Figure 4.1.

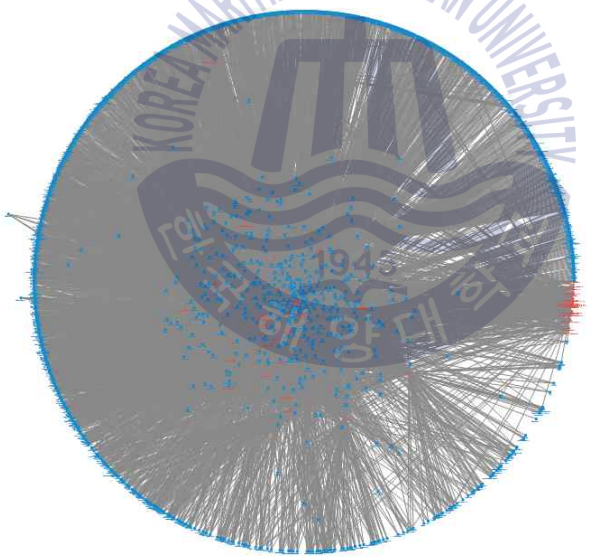


Fig. 4.1 Two mode visualization of keyword networks

18) This option is available by following Visualize >> Spring >> 2D >> Kamada and Kawai in order.

As seen in Figure 4.1, all the source nodes with a red color are located in the center and target nodes are located inside and outside of the network. Keyword networks construct structures with strong density. By following the similar procedure, I prepare two mode input data for collocation networks. The difference is that collocation networks use collocates as target nodes. By using Kamada-Kawai's spring algorithm, the two-mode visualization of collocation networks is shown in Figure 4.2.

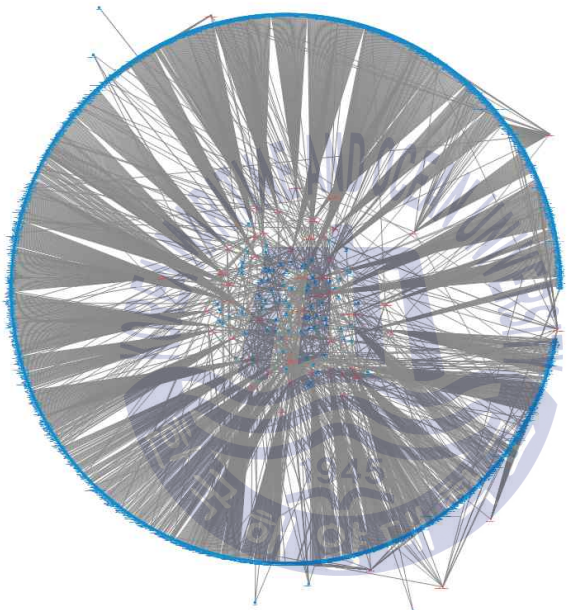


Fig. 4.2 Two mode visualization of collocation networks

As seen in Figure 4.2, the collocation network structure is less dense than the keyword network structure, showing that the collocates are less connected to each other.

As a next step to create one mode networks, I use a transform menu in order to transform two mode keyword networks and collocation networks into one mode networks. For this purpose, I transform two mode data into one mode data by using one of proximity measures such as the cosine similarity in order to find statistically significant relationships between keywords.[19] The final output of one mode matrix for keyword networks is shown in Table 4.6.

Table 4.6 One mode data for keyword networks

Number	Sources	Targets	Weight
1	SHALL	RESCUE_BOATS	0.0
2	SHALL	REQUIREMENTS	0.1
3	SHALL	REQUIRED	0.1
4	SHALL	REGULATION	0.1
5	SHALL	PROVISIONS	0.1
.	.	.	.
25	SHALL	COMPETENT	0.1
1	BALLAST_WATER	STEERING_GEAR	0.3
2	BALLAST_WATER	RESCUE_BOATS	0.2
3	BALLAST_WATER	SURVIVAL_CRAFT	0.3
4	BALLAST_WATER	OIL_RECORD_BOOK	0.2
5	BALLAST_WATER	OIL_TANKERS	0.3
6	BALLAST_WATER	ON_BOARD	0.4
.	.	.	.
18	BALLAST_WATER	BULK_CARGOES	0.2
1	OIL_RECORD_BOOK	OIL_TANKERS	0.2
2	OIL_RECORD_BOOK	ON_BOARD	0.2
3	OIL_RECORD_BOOK	RESCUE_BOATS	0.2
4	OIL_RECORD_BOOK	STEERING_GEAR	0.2
5	OIL_RECORD_BOOK	SURVIVAL_CRAFT	0.2

19) In the NetMiner 4.0, I follow Transform >> Mode >> two-mode network to build one-mode data for keyword networks and collocation networks. After each one-mode data is created, I eliminated diagonal values which are self indication because these diagonal values always link to themselves. I followed Transform>>Value>>Diagonal.

As seen in Table 4.6, the weight column is added to the network data. This weight indicates the strength between keywords. In the same way, I created one mode data for collocation networks.

In order to reduce less significant relationships of keyword networks and collocation networks, I use the cosine similarity weight as a cut-off threshold which produces about 250 links between keywords. For visualization of keyword networks, 0.185 of cosine similarity weight is the most proper value producing 251 links. On the other hand, for visualization of collocation network, 0.004 of cosine similarity weight is the most proper value producing 248 links.

4.2.3 Centrality and Cohesion Algorithms

In order to know which node is more important than others, the notion of centrality is used because central nodes may be more important in the networks than peripheral nodes. There are three important centralities which have been widely used in social network analysis: degree centrality, closeness centrality, and betweenness centrality, as suggested by Freeman (1979). Degree centrality is used to measure local centrality, considering indegree and outdegree centralities. Closeness centrality is computed based on the global centrality, using path distances between nodes. Betweenness centrality is used to measure the degree of node positions between nodes. As the extended version of Freeman's concepts, eigenvector centrality was proposed by Bonacich (1987). Eigenvector centrality[20] assigns high centrality scores to the nodes that connect higher scoring nodes than low scoring

20) Google's PageRank is also a type of variant of the eigenvector centrality.

nodes. Because of this property, eigenvector centrality can reflect the degrees of both local and global centrality together. For language network analysis, I select two centrality measures such as eigenvector and betweenness centrality to find influential nodes in keyword networks and collocation networks. I analyse eigenvector centrality by using Netminer 4.0.[21] The algorithm is computed as follows:

$$(4) \qquad x_i = \frac{1}{\lambda} \sum_{j=1}^{N} a_{ij,} x_j$$

In (4), λ is a constant, N is the total number of nodes, a_{ij} calculates node connections, and x_j means eigenvector centrality value for node j. Eigenvector centrality has an advantage which counts the number of node words adjacent to a given node by giving weight each of the adjacent nodes by its centrality (Borgatti, 2013). On the other hand, betweenness centrality (Freeman, 1979) measures the extent to which a node lies between all the other pairs of nodes on their geodesic links.[22] The algorithm is as follows:

$$(5) \qquad b_{ij}(p_k) = \frac{g_{ij}(p_k)}{g_{ij}}$$

21) I follow Analyze > Centrality > eigenvector
22) It is computed by clicking Analyze >> Centrality >> Betweenness

In (5), p_k means a specific node, g_{ij} is the shortest path which links node *i* and node *j*, and $g_{ij}(p_k)$ indicates the number of shortest path including p_k among the shortest path of node *i* and node *j*.

4.3 Comparison of Keyword Networks and Collocation Networks

I compare keyword networks with collocation networks in order to know which networks are better to explain textural features in the ESP texts because textural features represent specific characteristics for a study corpus.

4.3.1 Centrality Structures: Eigenvector and Betweenness

I visualize the structures of keyword networks and collocation networks using two centrality measures such as eigenvector and betweenness. First of all, let us consider keyword network structures of eigenvector and betweenness. They are shown in Figure 4.3 and Figure 4.4 respectively.

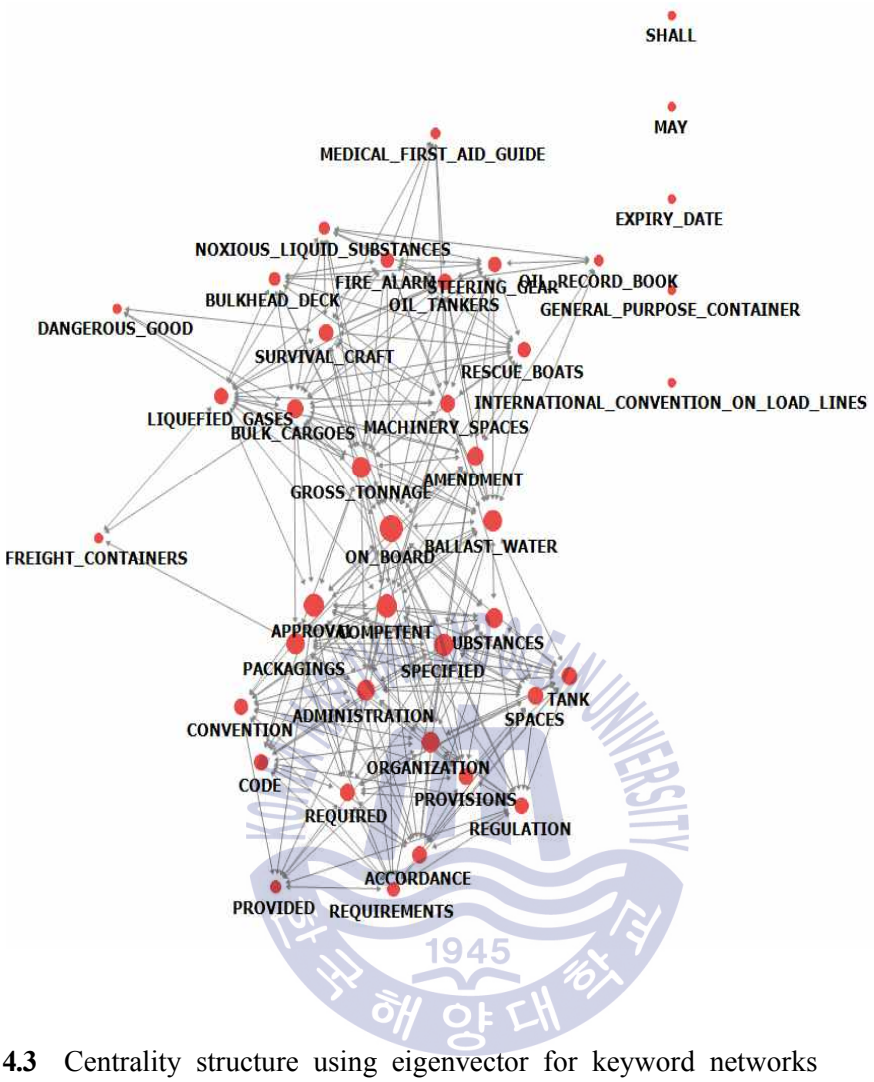


Fig. 4.3 Centrality structure using eigenvector for keyword networks

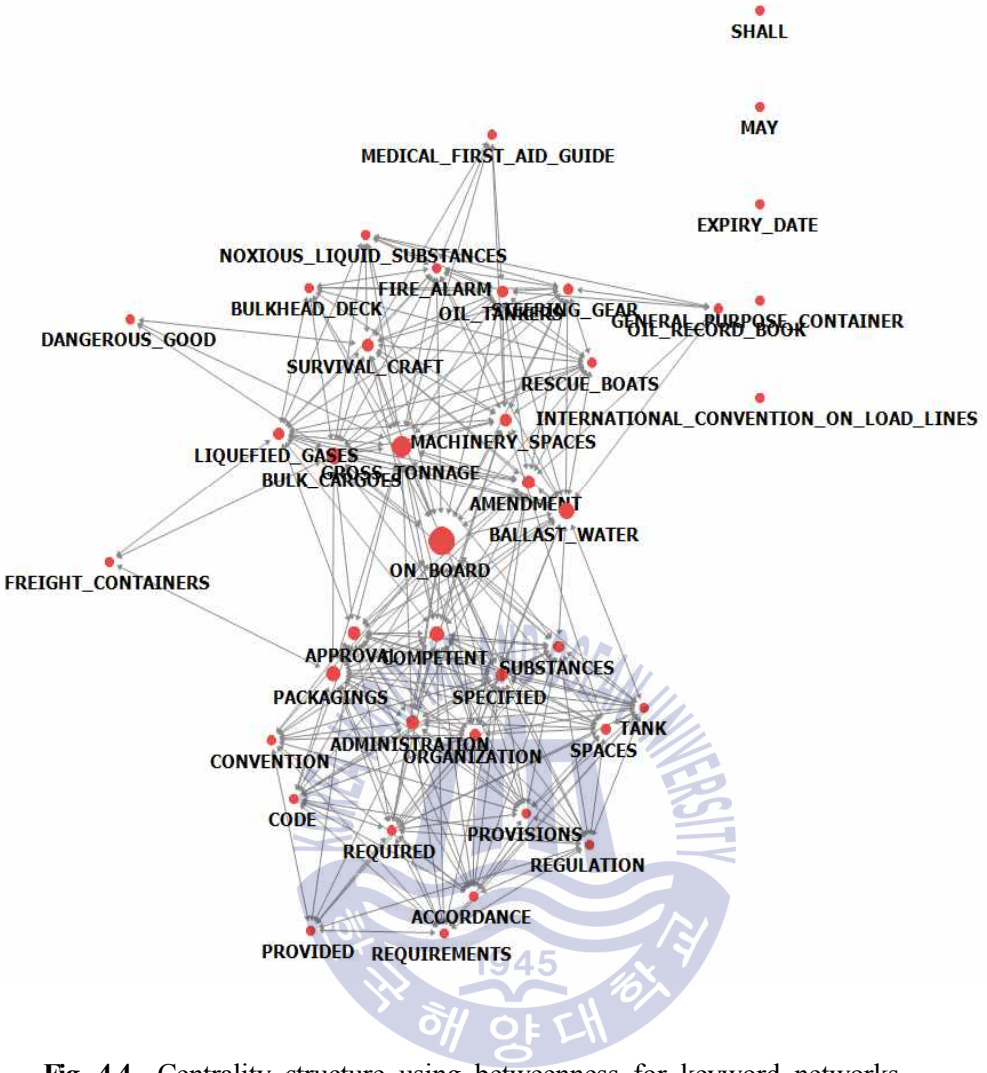


Fig. 4.4 Centrality structure using betweenness for keyword networks

As seen in Figure 4.3 and Figure 4.4, they might have the same structures at a glance. Because eigenvector centrality and betweenness centrality produce quite different ranks among keywords, is it the same structures that they have? To answer this question, we need to compare any nodes which show quite different ranks between eigenvector centrality and betweenness centrality.

In order to explain structural differences of eigenvector and betweenness centrality in keyword networks, I selected an example of *survival_craft* which comes 18th with 0.143 of eigenvector value but comes 12th with 0.015 of betweenness value. Surprisingly, the structures of *survival_craft* in eigenvector and betweenness centrality in keyword networks are really different, as shown in Figure 4.5 and Figure 4.6 respectively.

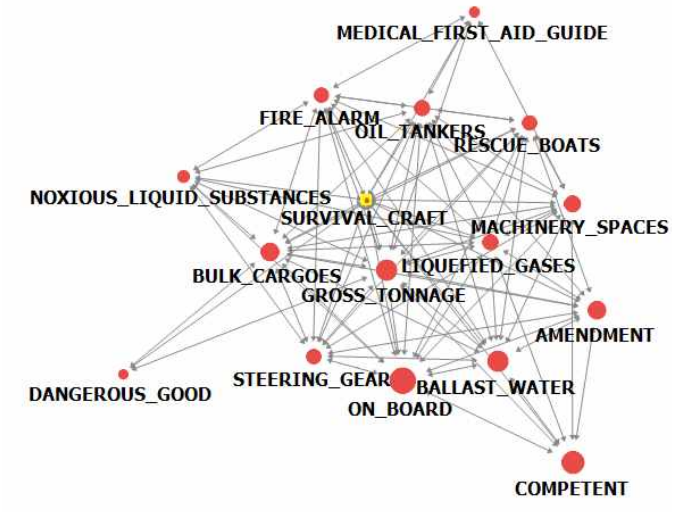


Fig. 4.5 Visualization of *survival_craft* using eigenvector in keyword networks

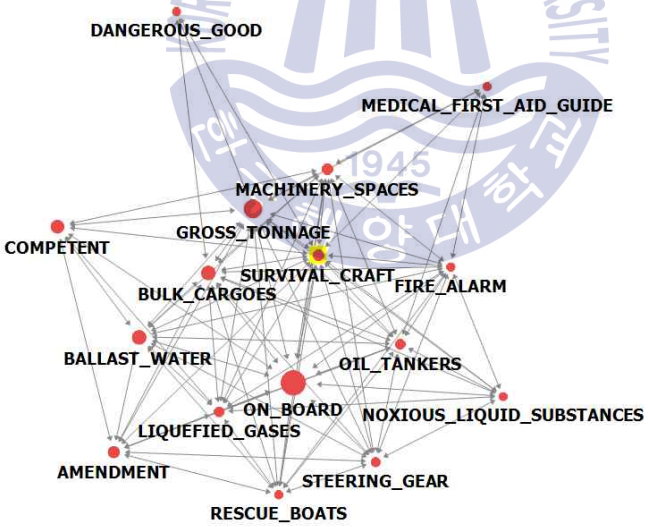


Fig. 4.6 Visualization of *survival_craft* using betweenness in keyword networks

As seen in Figure 4.5 and 4.6, the eigenvector structure for *survival_craft* shows that it is a small node and does not have many links to its neighbors, whereas the betweenness structure for *survival_craft* indicates that it is a bigger circle node. This finding shows that the structures of eigenvector centrality and betweenness centrality are really different, although they might superficially look like the same structures in Figure 4.4 and Figure 4.5. Now, let us consider collocation network structures of eigenvector and betweenness. They are shown in Figure 4.7 and 4.8 respectively.

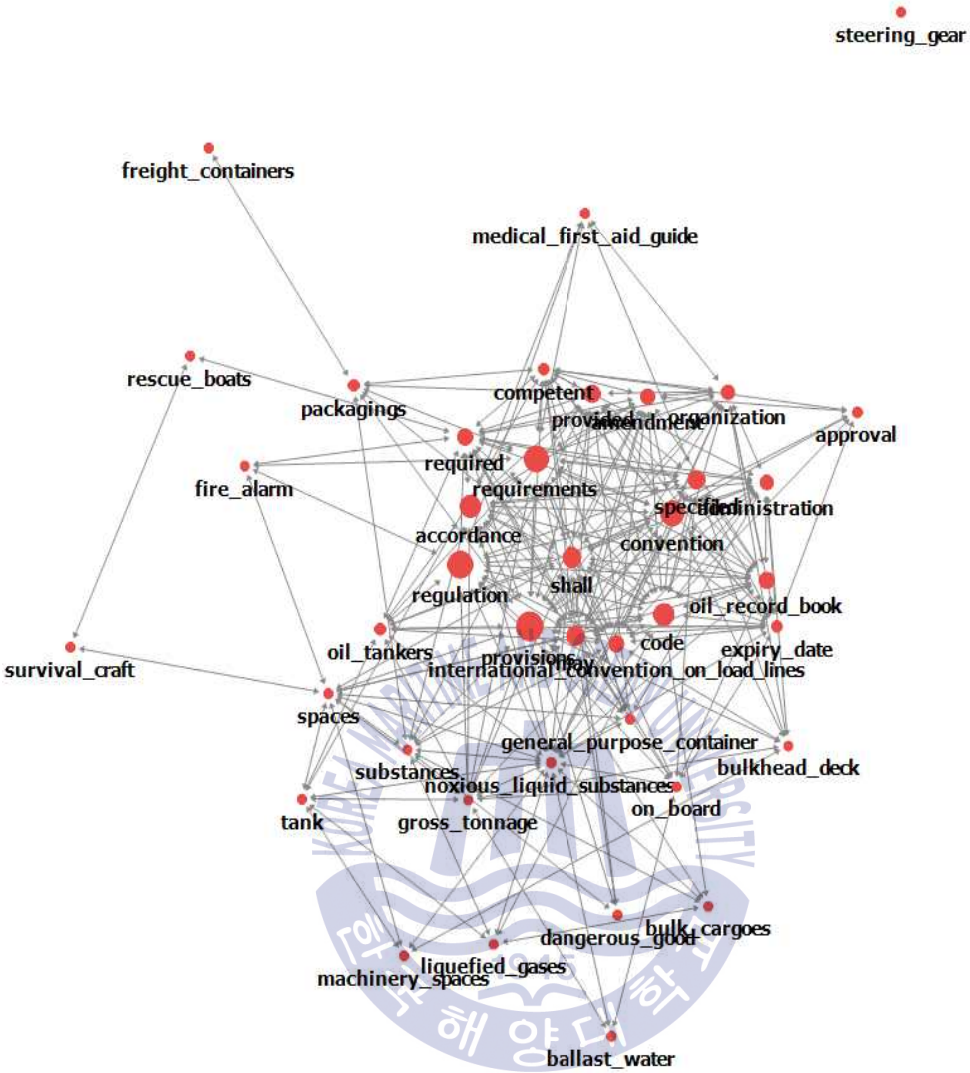


Fig. 4.7 Centrality structure using eigenvector for collocation networks

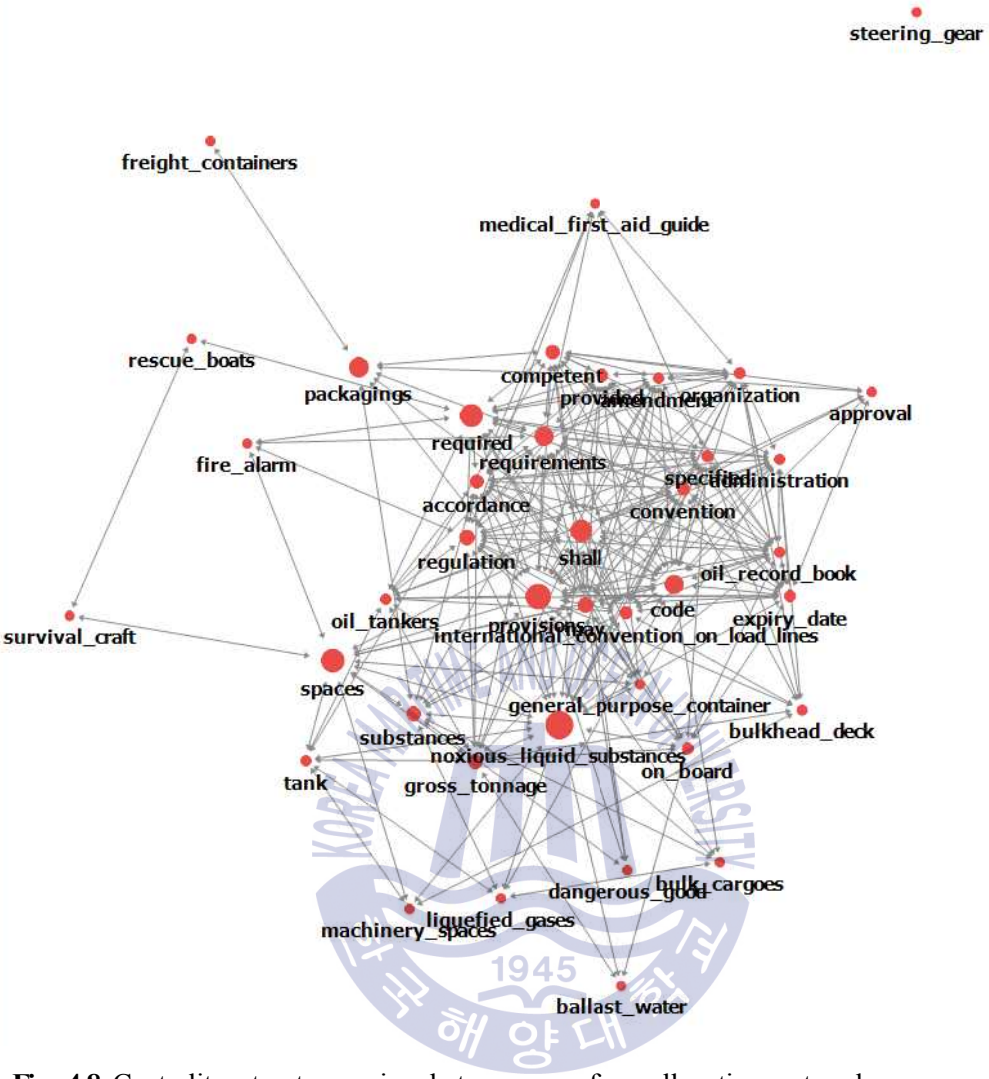


Fig. 4.8 Centrality structure using betweenness for collocation networks

As seen in Figure 4.7 and Figure 4.8, they might have the same structures at a glance as discussed in keyword network structures. In the same vein, we need to compare any nodes which show quite different ranks between eigenvector centrality and betweenness centrality in collocation networks.

In order to explain structural differences between eigenvector and betweenness centrality in collocation networks because these two structures look like the same, I selected *noxious_liquid_substances* which comes 21th with 0.037 eigenvector value but comes 1th with 0.089 betweenness value. The structures of *noxious_liquid_substances* in eigenvector and betweenness centrality in collocation networks are really different, as shown in Figure 4.9 and 4.10 respectively.

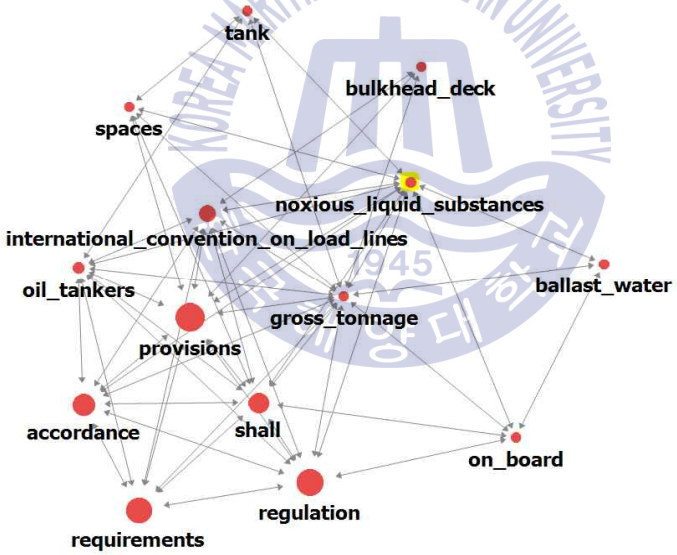


Fig. 4.9 Visualization of *noxious_liquid_substances* using eigenvector in collocation networks

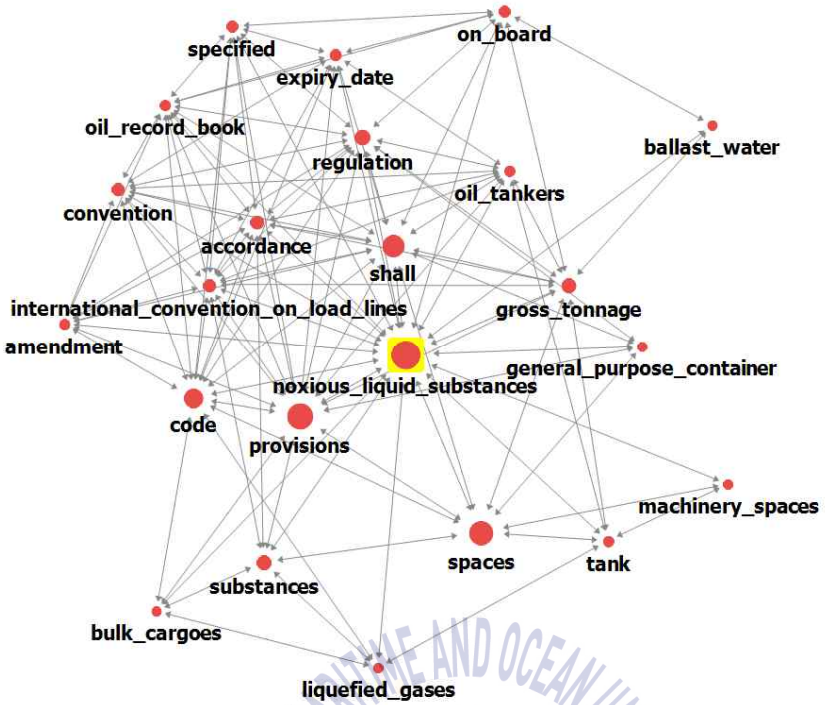


Fig. 4.10 Visualization of *noxious_liquid_substances* using betweenness in collocation networks

As seen in Figure 4.9 and Figure 4.10, the eigenvector structure for *noxious_liquid_substances* shows that it is a small node and does not have many links to its neighbors, whereas the betweenness structure for *noxious_liquid_substances* indicates that it is located at the center and the size of the node is the largest in all the other nodes. This result shows that the structures of eigenvector centrality and betweenness centrality in collocation networks are really different, as discussed in keyword network structures.

Based on the structural difference between keyword networks and collocation networks, we need to investigate which networks are better to explain textural features represent specific characteristics for a study corpus. For the first experiment, we need to a sample of specific purpose terms which are listed on a reference multi-word compound list and general purpose terms which are not. Therefore, I use a list of 40 vocabulary items consisting of 20 single words and 20 multi-word compounds which are source keywords for network analysis in a study corpus discussed in Table 4.2 in section 4.2.1. Experiment results of centrality analysis are shown in Table 4.7.

Table 4.7 Top 20 multi-word specific purpose terms versus top 20 general purpose terms

(number/%)

Network Structures		Keyword Networks		Collocation Networks	
		Specific Purpose Terms	General Purpose Terms	Specific Purpose Terms	General Purpose Terms
Centrality	Eigenvector	7/35%	13/65%	4/20%	16/80%
	Betweenness	9/45%	11/55%	4/20%	16/80%

As seen in Table 4.7, in the keyword networks, eigenvector centrality showed similar percentages such as 7 specific purpose terms (35%) and 13 general purpose terms (65%). Betweenness centrality showed 9 specific

purpose terms (45%) and 11 general purpose terms (55%). There are no big differences between them within keyword networks. In the collocation networks, however, eigenvector centrality showed very predominant general purpose terms such as 16 general purpose terms (80%) and 4 specific purpose terms (20%). Interestingly, betweenness centrality identically showed the same results.

For further exploration for the centrality of collocation networks, I present two ranking lists of keywords in Table 4.8, according to values of eigenvector centrality and betweenness centrality. The top 20 keywords are sorted based on the eigenvector value in the left column and the betweenness value in the right column. Special purpose terms are highlighted in the grey color. Each of the two centrality measures includes four different specific purpose terms.[23)]

23) Some of the specific purpose terms can appear on a list of general purpose terms because these four are likely to link more collocates of general purpose terms than other specific purpose terms do (personal communication with my advisor, Professor Jhang).

Table 4.8 Top 20 ranked keywords in collocation networks

Rank	Term Type	Keywords	*E-Value	Term Type	Keywords	**B-Value
1	General	provisions	0.407	Specific	noxious_liquid_substances	0.090
2	General	regulation	0.386	General	provisions	0.079
3	General	requirements	0.351	General	spaces	0.068
4	General	convention	0.337	General	required	0.064
5	General	accordance	0.285	General	shall	0.059
6	General	code	0.269	General	packagings	0.051
7	General	shall	0.230	General	code	0.049
8	General	may	0.206	General	requirements	0.047
9	General	specified	0.185	General	may	0.029
10	General	provided	0.180	General	regulation	0.027
11	General	required	0.161	General	competent	0.025
12	Specific	oil_record_book	0.151	Specific	gross_tonnage	0.024
13	Specific	international_convention_on_load_lines	0.146	General	substances	0.024
14	General	amendment	0.142	General	accordance	0.018
15	General	organization	0.118	Specific	international_convention_on_load_lines	0.016
16	General	administration	0.117	General	convention	0.016
17	Specific	oil_tankers	0.058	General	provided	0.014
18	Specific	expiry_date	0.052	General	specified	0.013
19	General	competent	0.051	Specific	on_board	0.012
20	General	packagings	0.042	General	organization	0.012

*E-value means eigenvector value. **B value means betweenness value.

In addition, the four different general purpose terms do not appear within the top 20 keywords sorted by the value of eigenvector and betweenness centrality, as listed in Table 4.9.

Table 4.9 General purpose terms under top 20 in centrality ranking

Term Type	Keywords	E-Value	Term Type	Keywords	B-Value
General	approval	0.0323	General	tank	0.008
	spaces	0.0177		amendment	0.006
	substances	0.0166		administration	0.005
	tank	0.0029		approval	0.000

It is expected that these general purpose terms such as *approval, spaces, substances, tank, amendment,* and *administration* should be included in the top 20 ranked keywords. Contrary to our expectation, they do not appear in collocation networks. It is hard to answer the reason right now, but it can be noteworthy that collocation networks using both eigenvector centrality and betweenness centrality seem to be more useful to find important general purpose items than keyword networks.

4.3.2 Cohesion Structures: Eigenvector and Betweenness

To investigate cohesion structures of keyword networks and collocation networks, I used community eigenvector and community betweenness measures. The results show that cohesion analysis is effective only to keyword networks because specific purpose terms and general purpose terms

are identified in the two giant groups in a community structure using eigenvector for keyword networks, whereas community betweenness fails to group the two term types. In keyword network cohesion structures, community eigenvector showed the better results than community betweenness, producing the two large groups more clearly, as seen in Figure 4.11 and 4.12, which show community structures using eigenvector and betweenness respectively.

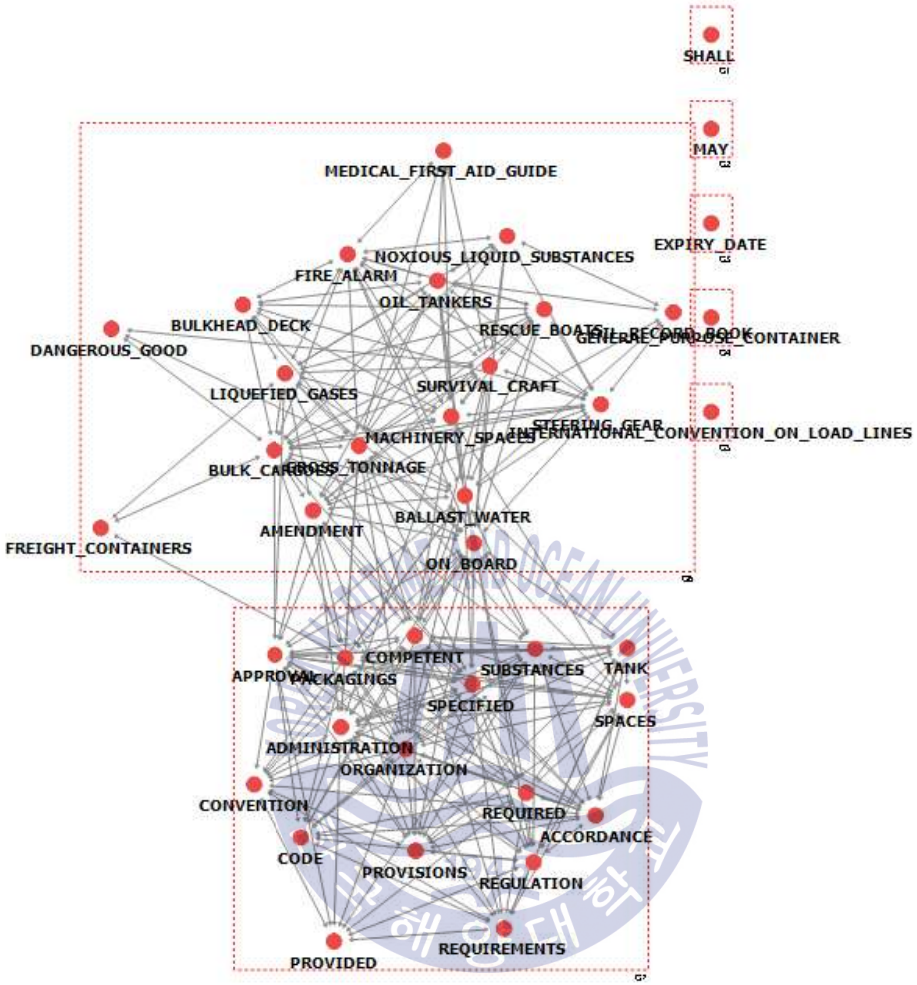


Fig. 4.11 Community structure using eigenvector for keyword networks

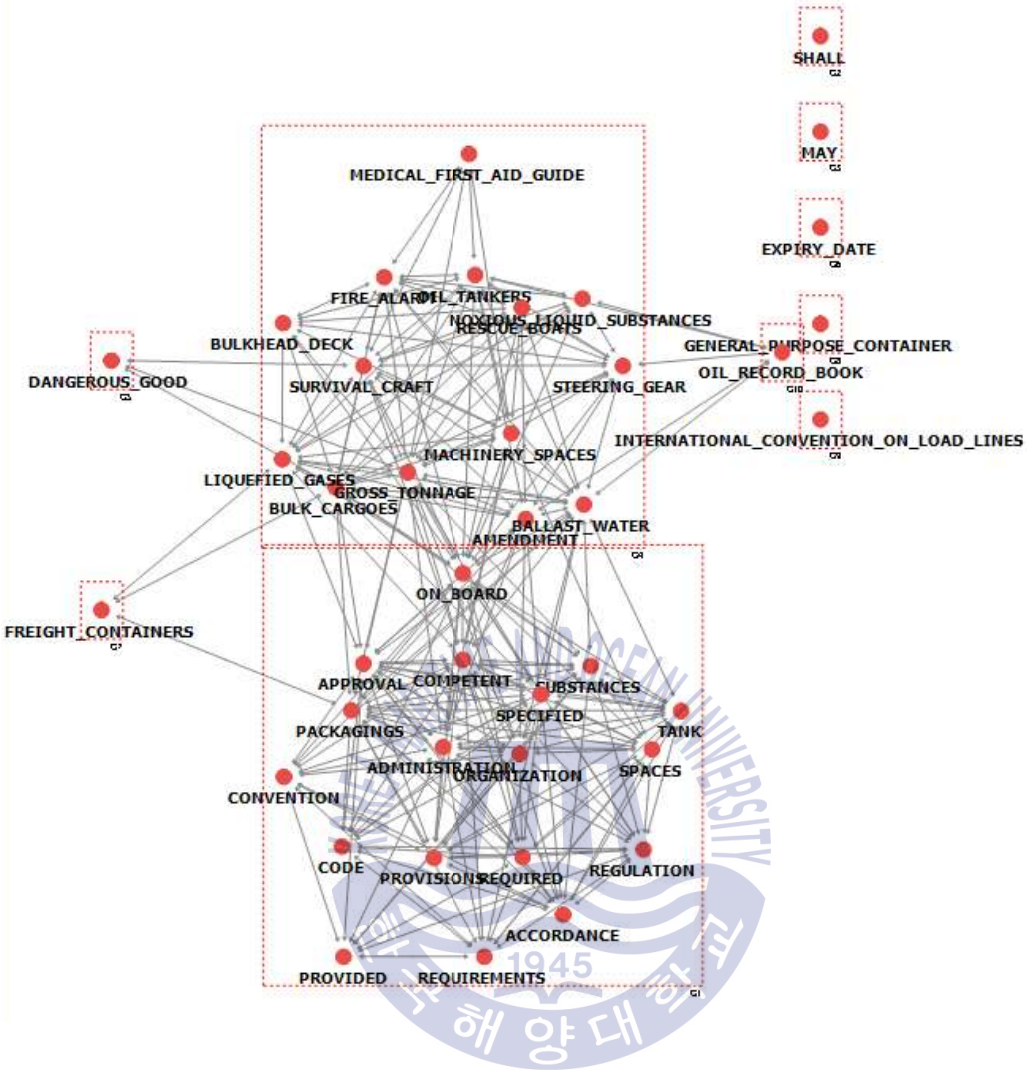


Fig. 4.12 Community structure using betweenness for keyword networks

Figure 4.11 and Figure 4.12 above look very similar in that they have two large communities. The difference is the number of small groups in that the eigenvector community structure has five small groups but the betweenness community structure has seven small groups.

Using collocation networks, community structures using eigenvector and betweenness for collocation networks are shown in Figure 4.13 and Figure 4.14 respectively.

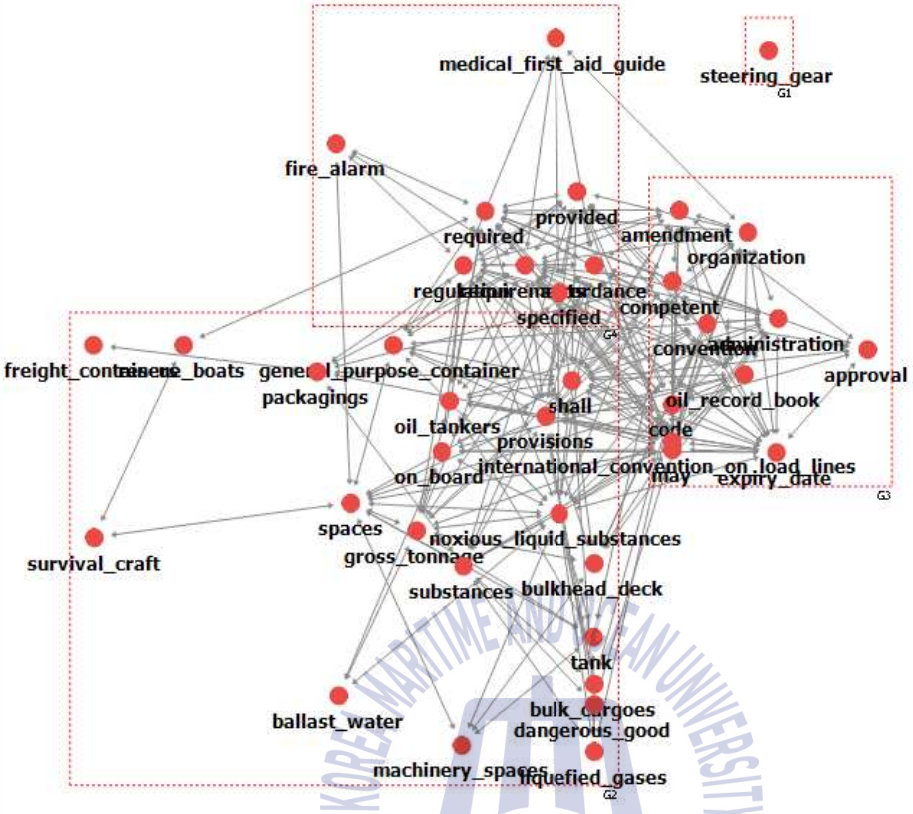


Fig. 4.13 Community structure using eigenvector for collocation networks

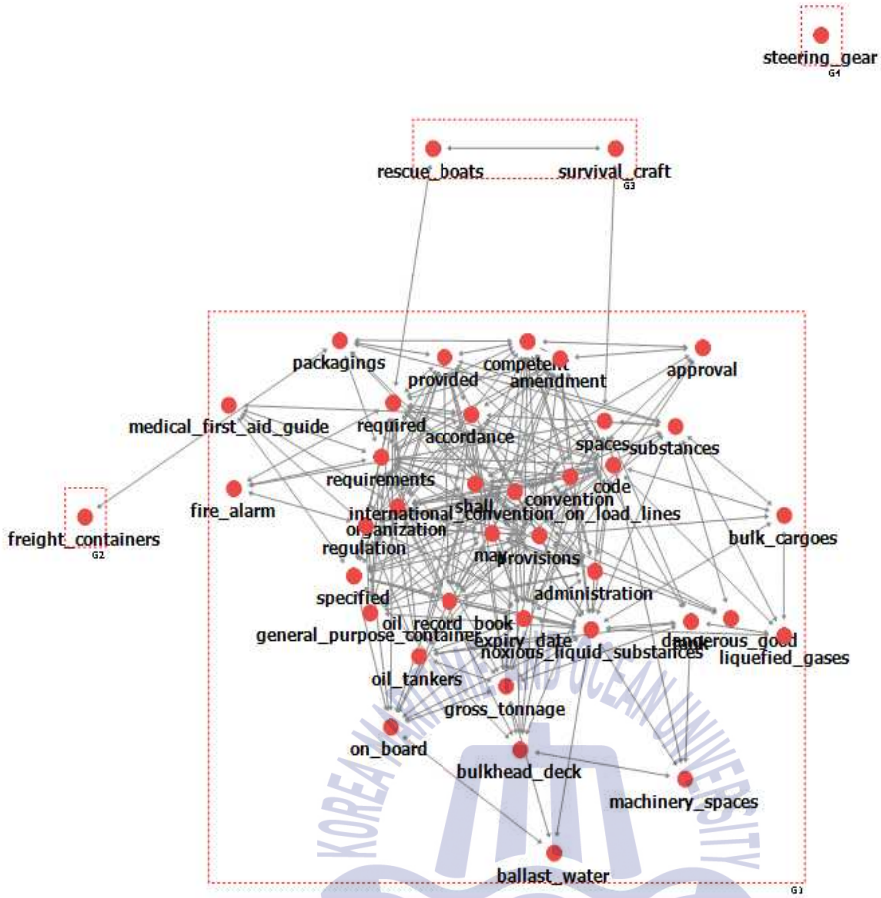


Fig. 4.14 Community structure using betweenness for collocation networks

The structure of Figure 4.13 shows three medium size communities, whereas that of 4.14 shows only one large community. Based on these observations, I find that keyword networks are interesting because they can divide two large community groups. Therefore, I focus on keyword networks because they show the better classification of communities than collocation networks to give a further explanation of community structures and provide pedagogical implications of the results. I analyzed the two large groups in keyword networks. Interestingly, each group prefers either specific purpose terms or general purpose terms. For example, Group 6 has 17 specific purpose terms (85%), whereas Group 7 has 17 general purpose terms (85%), as shown in Table 4.10. This finding implies that community eigenvector can be used to find specific purpose terms and general purpose terms among maritime English vocabulary items.

Table 4.10 Community structures with top 20 keywords for keyword networks

(number/%)

Network Structures		Keyword Networks			
		Specific Purpose Terms		General Purpose Terms	
Cohesion	Community Eigenvector	KG 6	17/85%	KG 6	1/5%
		KG 7	0/0%	KG 7	17/85%
	Community Betweenness	KG 1	1/5%	KG 1	17/85%
		KG 4	14/70%	KG 4	0/0%

However, it is necessary to critically observe these results by asking the following two questions: (i) What makes specific and general purpose terms belong to a different group? (ii) Why Groups 6 and 7 fails to capture all of the specific purpose terms or all of the general purpose terms? In order to answer these questions, I present some statistical information to show detailed attributes of each keyword in Groups 6 and 7, as shown in Table 4.11.

Table 4.11 Community eigenvector for keyword networks

Group 6 (17 Keywords)				Group 7 (17 Keywords)			
Specific Purpose Terms	Keyness	Freq.	Term Speci ality	General Purpose Terms	Keyness	Freq.	Term Speci ality
BALLAST_WATER	1,116.47	686	0.10	CONVENTION	4,200.24	3,449	-0.81
ON_BOARD	746.97	718	-0.03	PROVISIONS	4,152.81	2,509	-0.92
GROSS_TONNAGE	579.97	335	0.27	ACCORDANCE	3,840.35	1,941	-1.00
MACHINERY_SPACES	545.50	208	0.60	PROVIDED	3,278.91	2,321	-1.00
BULK_CARGOES	403.84	177	1.00	SPACES	3,018.15	1,285	-0.86
SURVIVAL_CRAFT	304.98	125	0.76	ORGANIZATION	2,952.35	2,016	-0.84
OIL_TANKERS	308.97	198	0.93	REGULATION	2,892.48	1,985	-0.92
LIQUEFIED_GASES	257.82	92	0.64	REQUIREMENTS	2,837.68	2,208	-1.00
STEERING_GEAR	240.16	99	0.92	CODE	2,749.68	1,860	-0.92
BULKHEAD_DECK	221.39	79	0.85	PACKAGINGS	2,590.06	924	-0.73
FIRE_ALARM	170	71	1.00	ADMINISTRATION	2,466.88	1,505	-0.85
RESCUE_BOATS	183.48	76	1.00	SUBSTANCES	2,374.74	1,040	-0.76
FREIGHT_CONTAIN ERS	172.84	70	0.33	TANK	2,323.06	1,154	-0.86
NOXIOUS_LIQUID_ SUBSTANCES	124.3	50	1.00	REQUIRED	2,313.15	1,865	-0.93
OIL_RECORD_BOOK	63.2	32	0.75	SPECIFIED	2,143.00	1,073	-0.80
MEDICAL_FIRST_A ID_GUIDE	45.87	19	1.00	COMPETENT	1,858.96	897	-0.73
DANGEROUS_GOOD	36.43	13	1.00	APPROVAL	1,837.82	1,069	-0.60

The above table shows the two large groups. Group 6 consists mainly of specific purpose terms, whereas Group 7 consists mainly of general purpose terms. These terms are sorted based on keyness. To explain the reason why Group 6 strongly prefers specific purpose terms and Group 7 strongly prefers general purpose terms, I provided columns for term speciality. For making the term speciality, I devised an equation which can inform us of what types of other vocabulary items are more connected to each keyword. This calculation is based on the observation of a keyword and its neighbor linkage. Term speciality is designed to measure the strength of specificity or generality of each keyword. The equation is as follows:

$$(6)\quad T(w) = \frac{S(w) - G(w)}{N}$$

I call an equation in (6) term speciality because it is used to show which type of term is stronger. Here, S refers to the number of the specific purpose terms in neighbors of w. G refers to the number of the general purpose terms in neighbors of w. Finally, N means a total number of neighbors. The output range is from positive 1 to negative 1. Positive 1 indicates that a keyword is connected to all of the specific purpose terms, whereas negative 1 means that a keyword is linked to all of the general purpose terms. In order to calculate a term speciality, I visualized each keyword and its neighbor by activating "Visualize Neighbor" option to observe network structures. For example, a word, *oil_tankers* and its

neighbors are visualized, as shown in Figure 4.15.

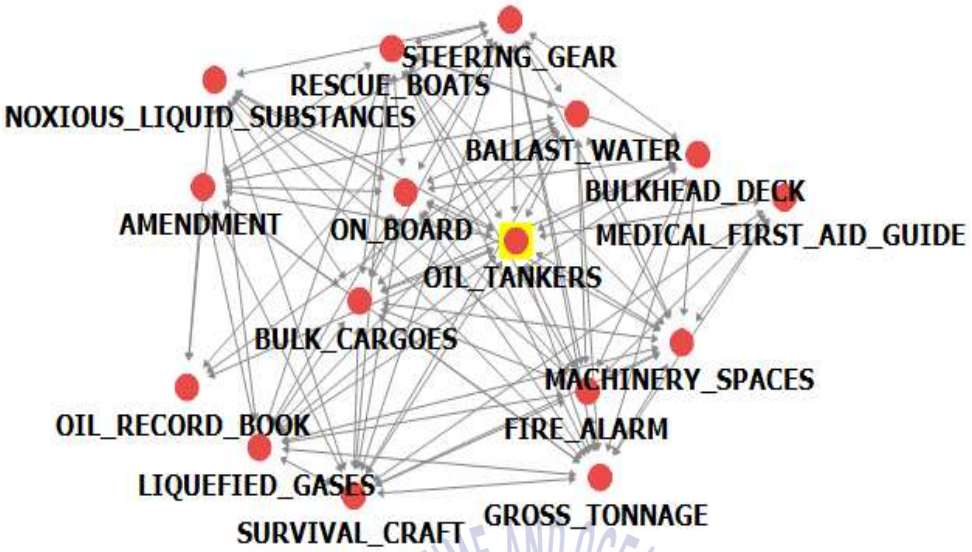


Fig. 4.15 Visualizing neighbors of *oil_tankers*

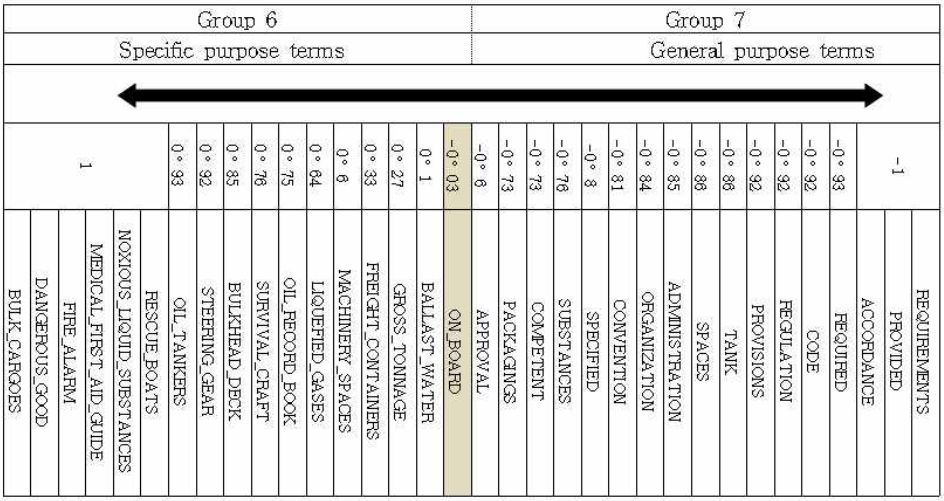
There are 14 multi-word compounds linked with *oil_takers*, whereas there is only one single word, *amendment*. Therefore, a term speciality is 0.93 indicating *oil_takers* has strong connections with specific purpose terms. I counted all the linked keywords of specific purpose terms and all the general purpose terms. I then calculated their term speciality.

The result of term speciality obtained in Table 4.11 above is likely to be strikingly straightforward. Most of the specific purpose terms have positive values ranging from 0.33 to 1. On the other hand, all general purpose terms have negative values ranging from -0.6 to -1. The result shows that general purpose terms are more strongly interconnected with

each other compared with specific purpose terms. Based on these results, I can make a hypothesis that specific purpose terms have positive term speciality, whereas general purpose terms have negative term speciality. So it is likely that eigenvector community analysis is an effective tool to identify specific purpose terms and general purpose terms in maritime English.

It implies that the specificity or generality in ESP texts seems to be not an absolute concept but close to the concept of the continuum which indicates strong or weak characteristics of specificity or generality. The similar previous study is Bhatia's (2009) research regarding English legal discourse which shows that legal vocabulary items are a continuum with either text-claritive function or text-cohering function. Therefore, I conclude that there are some degrees of specificity and generality in keyword network structures of Group 6 and Group 7, as shown in Table 4.12.

Table. 4.12 Textual clustering cohesion using community eigenvector



As seen in Table 4.12, most of the multi-word keywords which are located in Group 6 have positive term speciality values except *on_board*. The finding of *on_board* is quite interesting because it does not have a positive value but has negative 0.03 which indicates this keyword may be a general purpose term. To give a further explanation for this, I show its network structure, as shown in Figure 4.16.

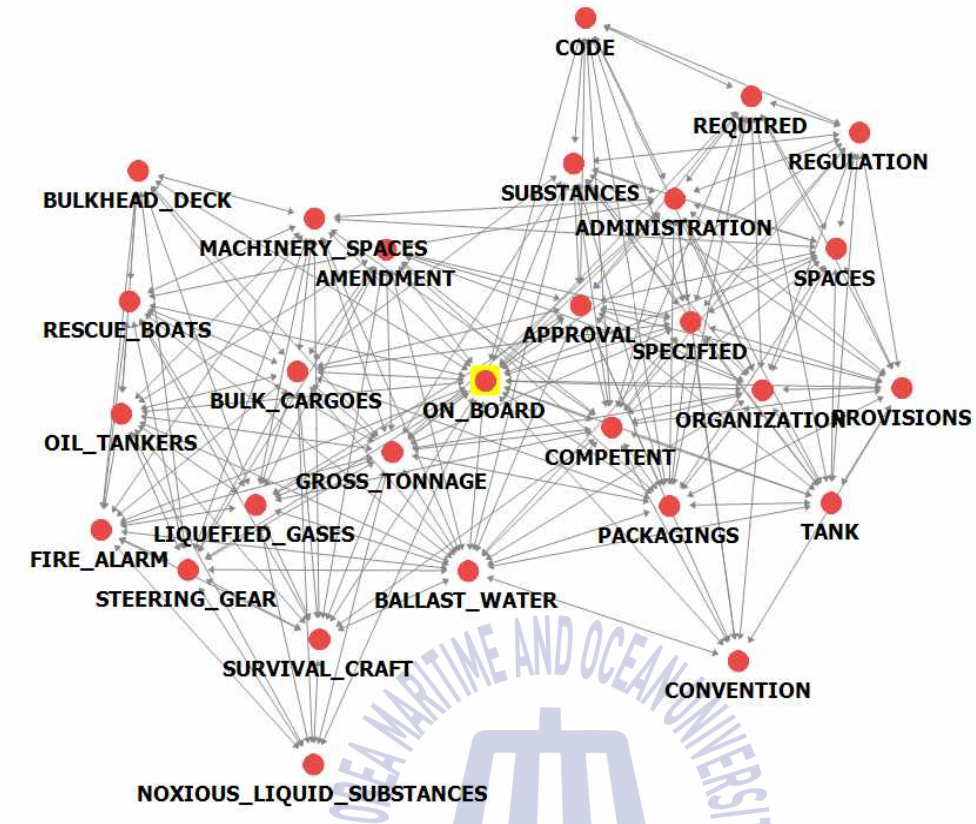


Fig. 4.16 Visualizing neighbors of *on_board*

As seen in Figure 4.16, *on_board* has 13 specific purpose terms and 14 general purpose terms. Thus, term speciality of *on_board* becomes negative 0.03.[24] This observation implies that the degree of a term speciality varies according to the keywords.[25] Term speciality shows that

24) Interestingly, amendment, is classified as a Group 6 which represents specific purpose terms. Amendment is not selected as a dictionary entry but it is classified as a specific group in my language network analysis. On the other hand, the entire keywords in Group 7 have all negative values.

the degree of specificity and generality is a continuum concept which is not absolute.[26]

Let us turn to the second question: Why Group 6 and Group 7 fail to capture all of the specific purpose terms or all of the general purpose terms? Keywords which Group 6 fails to capture are *expiry_date, general_purpose_container,* and *international_convention_on_load_lines*, whereas those which Group 7 fails to capture are *shall, may,* and *amendment*. Now, I look at small groups to give explanation of what makes these keywords separated from large groups. Table 4.13 shows some information on separated small keyword groups.

25) This finding can be an analogy of a linguistic phenomenon such as grammaticalization. Grammaticalization means a process of language change. For example, lexical categories such as nouns or verbs can be transformed to be different grammatical categories such as affixes or prepositions. Textual clustering cohesion using community eigenvector also implies a process of word change of meaning. This possibility is supported by various term speciality in the network structures discussed earlier. In the personal communications with my advisor, Professor Jhang, he told me that there is likely to be a possibility for some of the specific purposes terms can be changed into general purpose terms and vice versa in some reasons like grammaticalization or pragmaticalization. For example, on_board is considered to be a general purpose item even if it is listed in the maritime English dictionary entry, whereas amendment belongs to a group of specific purpose terms although it is not listed in the maritime English dictionary entry.

26) Bhatia (2009)'s English legal discourse study is in line with my results. This study categorized legal vocabulary items as a continuum with either text-claritive function (e.g., *falling within, specified*, and *set out*) or text-cohering function (e.g., *described, mentioned,* and *referred to*).

Table 4.13 Small groups of community eigenvector in keyword networks

Keywords	Keyness	Freq.	Group	Linked Keywords	Specific/ General	Term Speciality
SHALL	35,662.63	16,380	1	254	General	-1
MAY	2,388.07	4,085	2	255	General	-1
EXPIRY_DATE	214.54	84	3	12	Specific	+1
INTERNATIONAL_CONVENTION_ON_LOAD_LINES	133.61	51	5	11	Specific	+1
GENERAL_PURPOSE_CONTAINER	25.19	9	4	4	Specific	+1

As seen in Table 4.13, two keywords listed as general purpose terms appear in two different groups. For example, *shall* is in Group 1 and *may* in Group 2. It is assumed that these auxiliaries do not have enough linked keywords which can contribute to making relationships with other general purpose terms. On the other hand, three keywords listed as specific purpose terms also appear in three different groups. For example, *expiry_date* is in Group 3, *general_purpose_container* is in Group 4, and *international_convention_on_load_lines* is in Group 5. The common property of these three keywords is that they all have the small number of linked keywords not to be connected to Group 6. This implies that these three terms have some unique concepts compared to other specific purpose terms.

Furthermore, there is another general purpose item to discuss why it does not appear in Group 7. It is a keyword, *amendment*, whose term speciality is positive 0.25. It means that *amendment* is connected to more specific purpose terms rather than general purpose terms. In addition, as

the keyness of *amendment* is very high (1,490.36), it is evident that this keyword is very important in a maritime law corpus. In this sense, I assume that lexicographers for a maritime English dictionary did not include *amendment* in any maritime English dictionaries because they are inclined to take very narrow criteria which make maritime dictionary more specific. Based on these observations, I suggest that the term, *amendment*, should be considered to be one of specific dictionary entry.

4.4 Critical Evaluation

An additional experiment was conducted to critically evaluate my hypothesis. I used the top 20 single specific purpose terms and the top 20 single general purpose terms to create language networks. I substituted the top 20 single specific purpose terms for multi-word compounds in order to check up again whether or not this additional experiment complies with the result obtained from the previous experiment. More specifically, let us make a judgement on which measure, eigenvector or betweenness, is better to identify specific or general purpose terms in language centrality structures because the previous experiment showed the same result of 80% in Table 4.7. Furthermore, let us check up whether cohesion structures using an eigenvector community measure in keyword networks can also distinguish a group of specific purpose terms and a group of general purpose terms. I used cosine similarity weight as a cut-off threshold which produces about 250 links between keywords. For visualization of keyword networks, 0.224 of cosine similarity weight is the most proper value producing 249 links. On the other

hand, for visualization of collocation networks, 0.005 of cosine similarity weight is the most proper value producing 251 links. Table 4.14 shows the results.

Table 4.14 Centrality and cohesion of top 20 general purpose terms versus 20 single word specific purpose terms

(number/%)

Network Structures		Keyword Networks				Collocation Networks	
		Specific Purpose Terms		General Purpose Terms		Specific Purpose Terms	General Purpose Terms
Centrality	Eigenvector	6/30%		14/70%		4/20%	16/80%
	Betweenness	7/35%		13/65%		8/40%	12/60%
Cohesion	Community Eigenvector	KG6	3/15%	KG6	15/75%	N/A	
		KG7	16/80%	KG7	1/5%		
	Community Betweenness	KG3	10/50%	KG4	4/20%		

As seen in Table 4.14, eigenvector centrality of collocation networks showed very predominant general purpose terms such as 16 general purpose terms (80%) and 4 specific purpose terms (20%), whereas betweenness centrality of collocation networks does not show any clear effect on classification that of eigenvector centrality does. For cohesion structures, only community eigenvector divides two large groups for the distinction between specific purpose terms and general purpose terms. Interestingly enough, each group prefers either specific purpose terms or general purpose terms. For example,

Group 6 has 15 specific purpose terms (75%), whereas Group 7 has 16 general purpose terms (80%).

To provide a further explanation of these results, I investigated a term speciality value to each group, as shown in Table 4.15.

Table 4.15 Community eigenvector for keyword networks

Group 7 (16 Keywords)				Group 6 (15 Keywords)			
Specific Purpose Terms	Keyness	Freq.	Term Speciality	General Purpose Terms	Keyness	Freq.	Term Speciality
FLAMMABLE	980.06	388	0.3	CONVENTION	4,200.24	3,449	-0.25
DISCHARGE	969.30	603	0.14	PROVISIONS	4,152.81	2,509	-1
ACCEPTANCE	837.65	516	0.5	ACCORDANCE	3,840.35	1,941	-1
VENTILATION	804.34	338	0.36	SPACES	3,018.15	1,285	-0.09
STOWAGE	565.99	262	0.47	ORGANIZATION	2,952.35	2,016	-0.27
DECK	507.81	524	0.28	REGULATION	2,892.48	1,985	-1
VALVE	380.72	187	1	CODE	2,749.68	1,860	-0.80
LIFERAFT	375.24	153	1	ADMINISTRATION	2,466.88	1,505	-0.60
MIXTURES	370.26	166	0.69	SUBSTANCES	2,374.74	1,040	-0.28
CONTAINERS	358.04	557	0.23	TANK	2,323.06	1,154	-0.23
MASTER	357.45	431	0.26	PACKAGINGS	2,590.06	924	0.36
LEAKAGE	336.83	164	0.84	REQUIRED	2,313.15	1,865	-1
BULKHEAD	324.75	147	1	SPECIFIED	2,143	1,073	-0.10
WATERTIGHT	276.73	122	1	COMPETENT	1,858.96	897	-0.10
WATERLINE	266.05	136	1	APPROVAL	1,837.82	1,069	0.00
BALLAST	237.13	201	0.75				

The result of term speciality shown in Table 4.15 is somewhat straightforward. Most of the general purpose terms have negative values ranging from 0.1 to 1 except one word, *packagings* showing positive 0.36. On the other hand, all the specific purpose terms have positive values ranging from 0.14 to 1. These results verify that a term speciality can also be used to explain why each group prefers either specific purpose terms or general purpose terms. However, there are some exceptions as well. A keyword, *amendment*, whose term speciality is positive 0.88, is located in Group 7. On the contrary, I found out that three single items such as *annex* (-0.08), *conditions* (-0.11), and *materials* (-0.15), whose term type are specific purpose terms, are classified in Group 6.

The rest of the small groups identified are shown in Table 4.16.

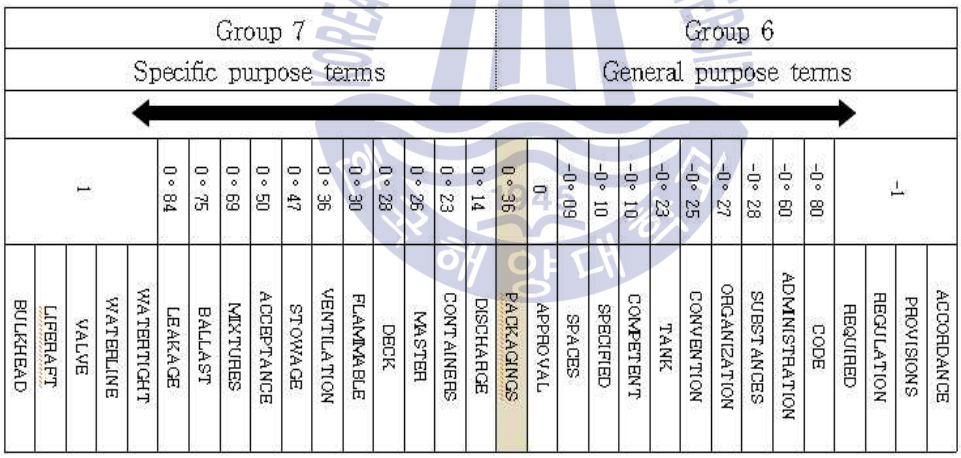
Table 4.16 Small groups of community eigenvector in keyword networks

Keywords	Keyness	Frequency	Group	Linked Keywords	Specific/ General	Term Speciality
SHALL	35,662.63	16,380	1	254	General	-1
MAY	2,388.07	4,085	4	255	General	-1
REQUIREMENTS	2,837.68	2,208	2	255	General	-1
PROVIDED	3,278.91	2,321	3	254	General	-1
SHIP	2,549.00	4,790	5	254	Specific	+1

As seen in Table 4.16, four keywords listed as general purpose terms appear in four different groups. It is thought that these words do not have enough linked keywords which can contribute to making relationships with other general purpose terms. On the other hand, one keyword listed as specific purpose terms also appears in a different group. For example, a keyword, *ship*, does not belong to Group 7 because its linked keywords do not have enough relationships for *ship* to link with other keywords.

This result conducted by my critical evaluation can be interpreted as the concept of the continuum which indicates strong or weak characteristics of specificity or generality, as shown in Table 4.17.

Table. 4.17 Textual clustering cohesion using community eigenvector



As shown in Table 4.17, the critical evaluation by the top 40 single word experiment with the top 20 specific purpose terms and the top 20 general purpose terms supports the result obtained from the previous experiment. Therefore, I conclude that the specificity or generality in ESP texts can be identified by keyword network structures regardless of either single words or multi-word compounds.

4.5 Summary and Implications

I summarize my findings discussed in Chapter 4. There are five findings obtained from the network structures. First, the structures created by centrality measures using eigenvector and betweenness showed that collocation network structures have more advantages over keyword network structures to find general purpose terms.

Second, the structures created by cohesion community using eigenvector and betweenness showed that keyword network structures have more advantages over collocation network structures to distinguish the two large communities belonging to specific or general purpose terms.

Third, my critical evaluation conducted in the top 40 single word experiment clarified and confirmed that eigenvector centrality is somewhat better than betweenness centrality in identifying general purpose terms. Fourth, it also clarified and confirmed that eigenvector community is somewhat better than betweenness community in identifying specific purpose terms through the same critical evaluation.

Finally, I made a term speciality equation to explain why a certain community can be separated into two groups for specific or general purpose terms, showing that the degree of specificity and generality is a continuum concept which is not absolute.

The results of this chapter can be used to provide pedagogical implications in creating ESP vocabulary lists. When teachers want to find maritime English vocabulary items for beginners or intermediate learners, they have to answer the question of what vocabulary they should teach. Using corpus linguistics methodology, teachers can make keyword lists in order to answer this question. Keyword lists show words with unusual frequencies as overused or underused words compared with general English. However, if there are too many keywords, how can teachers decide which keywords are more important than the others? Unfortunately, keyword analysis cannot provide rankings among keywords. Moreover, it is unsure that keywords are really key in texts, as noted in Scott (2015). However, language network analysis can be used to compensate for this explanatory gap. For instance, collocation network structures can be used to identify general purpose terms by using eigenvector centrality. In addition, cohesion structures of keyword networks tell us that a group of specific purpose terms and a group of general purpose terms are well separated by using eigenvector community.

Therefore, ESP teachers have advantages deciding what vocabulary to teach and how to classify vocabulary items in ESP through language

network analysis which I have proposed in this thesis.

Chapter 5. Conclusion

5.1 Summary

I summarize my research in this thesis. In Chapter 2, I explored maritime English as ESP. I discussed the most important two concepts such as keyness and collocation. Keyness was explored by its previous studies, strategies for deciding a reference corpus, the statistical measures for keyword analysis, and the problems of previous keyword analysis. Collocation was introduced through three types of collocations, the statistical measures for window collocations, and the problems of previous collocation analysis. This chapter introduced visualization studies in corpus linguistics. I reviewed visualization studies carried by several different perspectives such as text visualizations, and collocation networks. I then focused on some basic concepts for network analysis, previous studies using language networks, some important definitions used for this study and types of language network constructions. In Chapter 3, I discussed the method of compiling the MEC by presenting step-by-step procedures for making of four-million-word MEC that consists of academy, news, laws and textbooks genres. In addition, I built a reference multi-word compound entry items from the two maritime dictionaries and one supplement Web source for tagging compounds. Using these multi-word compounds, Python programs

were employed to create a new MEC which was tagged by special purpose multi-word compound items. In Chapter 4, I adopted language network analysis in order to provide further explanatory power to corpus linguistics data. I compared keyword networks with collocation networks using centrality and cohesion community based on eigenvector and betweenness. Using these methods, I conducted two experiments. The first experiment used 20 multi-word compounds and 20 single general purpose terms. The second experiment used 40 single words with half specific purpose terms and half general purpose terms. Finally, I suggested my implications of pedagogical aspects for maritime English education.

5.2 Findings and Implications

In Chapter 3, I found out a variety of differences by evaluating the MEC with and without multi-word compounds. There were many differences with regard to STTR, word lists, n-gram lists, and keyword lists. First, the STTR indicated a wider variety of vocabulary items in the multi-word compound tagged MEC. Second, the wordlist showed that there were more words in the multi-word compound tagged MEC than the untagged MEC. Third, the 4-grams with the tagged MEC contained fewer types than the untagged MEC. Fourth, there were more keywords in the multi-word compound tagged MEC than the untagged MEC. Almost all of the keywords from untagged and tagged corpora showed different dispersion plots. GraphColl 1.0 can visualize multi-word compound nodes and their linked collocates.

To provide a further explanatory power to corpus linguistics data, I adopted language network analysis. My network analysis and critical

evaluation led us to clarify and confirm that centrality structures created by eigenvector and betweenness in collocation networks have more advantages over keyword network structures to find general purpose terms. On the other hand, the cohesion community structures created by eigenvector and betweenness in keyword networks distinguish a group of the specific purpose terms and general purpose terms. More specifically, the eigenvector centrality structures in collocation networks represented better results than betweenness centrality in identifying general purpose terms. Contrastingly, the eigenvector cohesion community structures in keyword networks represented better results than betweenness in identifying specific purpose terms. I invented a term speciality algorithm to provide an explanatory power to the reason why keyword networks are separated into a large specific purpose term group and a large general purpose term group. Different degrees of term speciality values ranging from negative one to positive one indicate that the specificity or generality in ESP vocabulary items is not an absolute concept but close to the continuum concept.

There are two limitations in this thesis. The first limitation is to use a law genre, instead using the whole MEC. The reason is that there is no multi-word compound tagged general English corpus to make a keyword list of the MEC. If BNC or BNC Baby had had multi-word compounds, I would have used a general standard corpus as a reference corpus to extract keywords in maritime English. For not only my future research but also other ESP fields, I hope to see the advent of a general English corpus tagged by multi-word compounds. The second limitation is the size of my study corpus. Due to the huge amount of calculation load for language

network nodes, my research used a limited number of keywords through the sampling process.

There are several implications of corpus linguistics. My proposed method of tagging English multi-word compounds can help traditional concordance tools and corpus visualization tools to be developed to a new horizon beyond the current research because a corpus tagged with multi-word compounds can reflect authentic language usages. For pedagogical implications, ESP teachers have the advantage to decide what to teach in ESP vocabulary by using network analysis which I have proposed in this thesis. Language network analysis can help ESP researchers to identify general purpose items in collocation networks using an eigenvector centrality measure and specific purpose terms in keyword networks using an eigenvector community measure.

References

Alonso, A., Millon, C., & Williams, G. (2011). Collocational networks and their application to an e-Advanced learner's dictionary of verbs in science. In *Proceedings of New Applications for New Users* (pp.12-22). Electronic Lexicography in the 21st Century, Bled, Slovenia: Institue for Applied Slovene Studies.

Archer, D., & Culpeper, J. (2009). Identifying key sociophilological usage in plays and trial proceedings (1640-1760): An empirical approach via corpus annotation. *Journal of Historical Pragmatics*, *10*(2), 286-309.

Baker, P. (2006). *Using Corpora in Discourse Analysis.* London: Continuum.

Barabasi, A. L. (2002). *Linked: The New Science of Networks*. Cambridge: Perseus.

Beavan, D. (2008). Glimpses through the clouds: Collocates in a new light. In *Proceedings of Digital Humanities 2008* (pp.53-55), University of Oulu: Scottish Corpus of Texts and Speech.

Bednarek, M. (2012). Get us the hell out of here: Key words and trigrams in fictional television series. *International Journal of Corpus Linguistics, 17*(1). 35-63.

Bengio, Y., & Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. *Journal of Machine Learning Research*, *5*, 1089-1105.

Bertels, A., & Speelman, D. (2013). Keywords method versus calcul des spécificités: A comparison of tools and methods. *International Journal of Corpus Linguistics, 18*(4), 536-560.

Bhatia, V. (2009). Intertextual patterns in English legal discourse. In Belcher, D. (Ed.), *English for Specific Purposes in Theory and Practices* (pp.186-204). Ann Arbor, MI: University of Michigan Press.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ... : Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25*(3), 371-405.

Biber, D., & Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes, 26*(3), 263-286.

Bocanegra-Valle, A. (2012). Maritime English. In Chapelle, C. A. (Ed.), *The Encyclopedia of Applied Linguistics* (pp.3570-3583). Hoboken: Wiley-Blackwell.

Bonacich, P. (1987). Power and centrality: A family of measures. *The American Journal of Sociology*, *92*(5), 1170-1182.

Borgatti, S. B., Everett, M. G., & Johnson, J. C. (2013). *Analyzing Social Networks*. United Kingdom: Sage Publications.

Brezina, V., McEnery, T., & Wattam, S., (2015). Collocations in context. *International Journal of Corpus Linguistics, 20*(2), 139–173.

Chen, Y., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology*, *14*(2), 30-49.

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information and lexicography. *Computational Linguistics, 16*(1), 22-29.

Chung, E. G., & Kang, B. M. (2012). Occupation nouns and related words in newspaper articles. *The Linguistic Association of Korea Journal, 20*(2), 145-170.

Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes, 23*, 393-423.

Culpeper, J. (2002). Computers, language and characterisation: An analysis of six characters in Romeo and Juliet. In M-M. Ulla, O. Carin, & K. Merja (Eds.), *Proceedings of Conversation in Life and in Literature* (pp.11-30). ASLA Symposium, *15,* Uppsala, Universitetstryckeriet: Association Suedoise de Linguistique Appliquee

Culpeper, J. (2009). Words, parts-of-speech and semantic categories in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics, 14*(1), 29-59.

Cyram, (2015). *Netminer 4.0*. Seoul: Cyram Co. Ltd.

Dandapat, S., Mitra, P., & Sarkar, S. (2006). Statistical investigation of Bengali noun-verb(N-V) collocations as multi-word expressions. In *Proceedings of MSPIL* (pp.230-233), Mumbai, Asian Federation of Natural Language Processing.

Demmen, J. (2009). *Charmed and chattering tongues: investigating the functions and effects of key word clusters in the dialogue of Shakespeare's female characters*. M.A. dissertation. Lancaster: Lancaster University.

Dunning, T. (1993). Accurate methods for the statistics of surprise and Coincidence. *Computational Linguistics, 19*(1), 61-74.

Evert, S. (2004). *The Statistics of Word Cooccurrences: Word pairs and Collocations*. Ph.D. dissertation, Stuttgart, Germany: University of Stuttgart.

Faber, P., León Araúz, P. & Reimerink, A. (2014) Representing environmental knowledge in EcoLexicon. In E. Bárcena, T. Read, & J. Arus (Eds.), *Languages for Specific Purposes in the Digital Era* (pp.267-301). Gewerbestrasse, Switzerland: Springer International Publishing.

Firth, J. R. (1957). A synopsis of linguistic theory 1930-1955. In Palmer, F. (Ed.), *Selected Papers of J. R. Firth 1952-1959*. London: Longman. 168-205.

Fischer-Starcke, B. (2009). Keywords and frequent phrases of Jane Austen's Pride and Prejudice: A corpus-stylistic analysis. *International Journal of Corpus Linguistics, 14*, 492-523.

Francis, G. (1993). A corpus-driven approach to grammar: Principles, methods and examples. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and Technology* (pp.137-156). Amsterdam: John Benjamins.

Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks, 1*, 215-239.

Glossary of Nautical Terms. (2015). *In Wikipedia*. Retrieved February 1, 2015, from https://en.wikipedia.org/wiki/Glossary_of_nautical_terms.

Grabowsky, L. (2015). Keywords and lexical bundles within English pharmaceutical discourse: A corpus-driven description. *English for Specific Purposes, 38*, 28-33

Gries. S. T. (2013). 50-something years of work on collocations: What is or should be next ..., *International Journal of Corpus Linguistics, 18*(1), 137-165.

Halliday, M. A. K. (1961). Categories of a theory of grammar. *Word*, *17*(3). 241-292.

Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning, 24*, 287-298

Hong, S. C., & Jhang, S. E. (2010). The compilation of a maritime English corpus for ESP learners. *Korean Journal of English Language and Linguistics, 10*(4), 963-985.

Hudson, R. (2007). *Hudson Language Networks. The New Word Grammar*. United Kingdom: Oxford University Press.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*, 4-21.

IMO. (2009). *Maritime English*, *Model Course 3.17*. Albert Embankment: IMO.

Jhang, S. E., & Lee, S. M. (2012). Key clusters analyses of lexical bundles used in English for academic purposes: The biomed corpus. *The Linguistic Association of Korea Journal, 20*(4), 219-239.

Jhang S. E., & Lee, S. M. (2013a). Visualization of collocational networks: Maritime English keywords. *Language Research, 49*(3), 781-802.

Jhang, S. E., & Lee, S. M. (2013b). Clusters and key clusters in the Maritime English Corpus. *Journal of Language Sciences, 20*(4), 199-219.

Jhang, S. E., & Lee, S. M. (2014). Collocational patterns of maritime English vocabulary from a statistical point of view. *Journal of Language Sciences, 21*(1), 173-193.

Jones, K. S. (1971). *Automatic Keyword Classification for Information Retrieval*, London: Butterworths.

Kerchove, R. (1961). *International Maritime Dictionary*. Princeton: David Van Nostrand.

Kunchukuttan, A., & Damani, O. P. (2008). A system for compound noun multiword expression extraction for Hindi. In *Proceedings of 6th International Conference on National Language Processing* (pp.20-29), A Digital Archive of Research Papers in Computaional Linguistics Conference. Nagoya, Japan: Association for Computational Linguistics.

Le, T. N. P., & Harrington, M. (2015). Phraseology used to comment on results in the discussion section of applied linguistics quantitative research articles. *English for Specific Purposes*, *39*, 45-61.

Lee, S. S. (2012). *Network Analysis Methods*. Seoul: Non-hyung.

Lee, S. S. (2014). A content analysis of journal articles using the language network analysis methods. *Journal of the Korean Society for Information Management, 31*, 49-68.

Lee, Y. G., & Kang, B. M. (2012). The usage of trends of political keywords in newspapers. *Journal of Language Sciences, 19*(1), 203-234.

Liang, W., Shi, Y., Tse, C.K., Liu, J, Wang, Y., & Cui, X. (2009). Comparison of co-occurrence networks of the Chinese and English languages. *Physica A, 388*, 4901-4909.

Lindquist, H. (2009). *Corpus Linguistics and the Description of English*. Edinburgh: Edinburgh University Press.

Mahlberg., M. (2007). Clusters, key clusters and local textual functions in Dickens. *Corpora, 2*(1), 1-31.

Mahesh & Sinha, K. (2011). Stepwise mining of multi-word expressions in Hindi. In *Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World* (pp.110-115), Association

for Computational Linguistics Conference, Stroudsburg, USA: Association for Computational Linguistics.

Margan, D., & Meštrović, A. (2015). LaNCoA: A Python Toolkit for Language Networks Construction and Analysis. In *Proceedings of IEEE 38th International Convention on Information and Communication Technology, Electronics and Microelectronics* (pp.1961-1966.), Opatija: Institute of Electrical and Electronics Engineers.

Masucci, A., & Rodgers, G. (2006). Network properties of written human language. *Physical Review E, 74*, 1-8.

Meunier, F., & Granger, S. (2008). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins.

McArthur, T. (1981). *Longman Lexicon of Contemporary English.* London: Longman.

McEnery, T. (2006). *Swearing in English: Bad Language, Purity and Power from 1586 to the Present*. Abington: Routledge.

Michelbacher, L., Evert, S. & Schutze, H. (2011). Asymmetry in corpus-derived and human word associations. *Corpus Linguistics and Linguistic Theory, 7*(2), 245-276.

Monroe J. W., & Stewart, R. (2005). *Dictionary of Maritime and Transportation Terms*. Pennsylvania: Cornell Maritime Press.

Moretti, F. (2013). *Distant Reading*. New York: Verso.

Newman, M. E. J. (2001). Scientific collaboration networks. Network construction and fundamental results. *Physical Review, 64,* 01613, 1-8

Newman, M. E. J. (2005). Power laws, pareto distributions and Zipf's law. *Contemporary physics*, *46*(5), 323-351.

Nooy, W., Mrvar, A., & Batagelj, V., (2008). *Exploratory Social Network Analysis with Pajek*. Cambridge: Cambridge University Press.

Oakes, M. P. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.

Pecina, P. (2008). Reference data for czech collocation extraction. In *Proceedings of the LREC workshop towards a shared task for MWEs*. Marrakech, Morocco: Langauge Resources and Evaluation Conference.

Perlmutter., D. M. (1980). Relational grammar. In E. A. Moravcsik, and J. R. Wirth (Eds.), *Syntax and semantics: Current approaches to syntax*, *13*, 195-229.

Plag, I., Braun, M. Lappe, S., & Schramm, M. (2009). *Introduction to English Linguistics*. Mouton de Gruyter: Berlin.

Ramisch, S., Caseli, H. M., Villavicencio, A., Finatto, M. J. B, & Machado, A. (2010). A hybrid approach for multiword eExpression identification. In T. Pardo, A. Branco, A. Klautau, R. Vieira, & V. Lima (Eds.), *Proceedings of Computational Processing of the Portuguese Language, 9th International Conference* (pp.65-74). 9th International Conference, PROPOR, Porto Alegre: Brazil, Springer.

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics, 13*(4), 519-549.

Rayson, P. & Garside, R. (2000). Comparing corpora using frequency profiling. In *Proceedings of the Workshop on Comparing Corpora* (pp.1-8), Hong Kong, Association for Computational Linguistics.

Rayson, P., Archer, D., Piao, S. L., & McEnery, T. (2004). The UCREL semantic analysis system. In *Proceedings of the workshop on beyond named entity recognition semantic labelling for NLP tasks in association with 4th International Conference on Language Resources and Evaluation* (pp.7-12), Lisbon, Portugal: Senseval.

Rayson, P. & Mariani, J. (2009). Visualising corpus linguistics. In M. Mahlberg, V. González-Díaz, & C. Smith (Eds.), *Proceedings of the Corpus Linguistics Conference Article* (pp.20-23), Liverpool, UK: Corpus Linguistics Conference.

Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. In K. Ajimer, & B. Altenberg (Eds.), *English Corpus Linguistics* (pp.128-143), Cambridge: Cambridge University Press.

Sag, I. A., Baldwin, T., Bond, F., Copestake A., & Flickinger. D. (2002). Multi-word expression: A pain in the neck for NLP. In A. Gelbukh (Ed.), *Proceedings of Conference on intelligent text processing and computational linguistics*. *Lecture notes in computer sciences* (pp. 1-15), 2276, Heidelberg: Springer.

Smadja, F. (1993). Retrieving collocations from text: Xtract. *Computational Linguistics*, *19*(1), 143-177.

Sinclair, J., Jones, S., & Daley, R. (2004). *English Collocation Studies: The OSTI Report,* In R. Krishnamurthy (Ed.), London: Continuum.

Scott, J. (2000). *Social Network Analysis: A Handbook*. London: SAGE Publications.

Scott, M. (199I). *Demystifying the Jabberwocky: A Research Narrative*. Ph.D. dissertation, Lancaster: Lancaster University.

Scott, M. (1997). PC Analysis of key words and key key words. *System, 25*(2), 233-245.

Scott, M. (2001). Mapping key words to problem and solution. In M. Scott, & G. Thompson (Eds.), *Patterns of Text: in Honour of Michael Hoey* (pp.109-127). Amsterdam: John Benjamins.

Scott, M. (2010). What Can Corpus Software Do. In A. O'Keeffe, & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp.136-151). London: Routledge Handbooks.

Scott, M. (2015). *WordSmith Tools Version 6.* Liverpool: Lexical Analysis Software.

Scott, M. & Tribble, C. (2006). *Textual Patterns: Keyword and Corpus Analysis in Language Education*. Amsterdam: John Benjamins.

Sole, R. V., Corominas, B., Valverde, S., & Steels, L. (2005). Language networks: Their structure, function and evolution. *Technical Report Santa Fe Institute Working Paper, 12*(42), 1-9.

Stuart, K. & Botella, A. (2009). Corpus linguistics, network analysis and co-occurrence Matrices. *International Journal of English Studies, Special Issue*, 1-28.

Stubbs, M. (2001). *Words and Phrases*. Oxford, UK: Blackwell.

Stubbs, M. (2010). Three concepts of keywords. In M. Bondi, & M. Scott (Eds.), *Keyness in Texts* (pp.21-42). Amsterdam: John Benjamins.

Tanmoy C. (2013). Identification of noun-noun collocations as multi-word expressions in Bengali corpus. *In proceedings of 9th Workshop on Multiword Expressions* (pp.64-72). Georgia, Atlanta: Association for Computational Linguistics.

Tanmoy C., Das, D., & Bandyopadhyay, S. (2011). Semantic clustering: An attempt to identify multiword expressions in Bengali. In *Proceedings of workshop on multiword expressions: from parsing and generation to the*

real world (pp.8-13). Workshop on Multiword Expressions. Portland, USA: Association for Computational Linguistics,

Thornbury, S. 2010. What can a corpus tell us about discourse? In A. O'Keeffe, & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp.270-287). New York, NY: Routledge Handbooks.

Trenker, P. (2000). Maritime English: An attempt at an imperfect definition. In *Proceedings of* S*econd Asian IMLA Workshop on Maritime English*. Dalian, China (pp.1-8): Dalian Maritime University.

Wagner, A. (2000). Enriching a lexical semantic net with selectional preferences by means of statistical corpus analysis. In *Proceedings of ECAI-2000 Workshop on Ontology Learning* (pp.37-42), Berlin: The European Chapter of the ACL.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature, 393*, 440-442.

Williams, G. (1998). Collocational networks: Interlocking patterns of lexis in a corpus of plant biology research articles. *International Journal of Corpus Linguistics, 3*(1), 151-171.

Williams, G. (2002). In search of representativity in specialised corpora: Categorisation through collocation. *International Journal of Corpus Linguistics, 7*(1), 43-64.

Williams, R. (1976). *Keywords: A Vocabulary of Culture and Society*. London, UK: Fontana.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice. *Applied Linguistics*, *21*(4), 463-489.

Zhang, W., Yoshida, T., Tang, X., & Ho, T. B. (2009). Improving effectiveness of mutual information for substantival multiword

expression extraction. *Expert Systems with Applications, 36*(8), 10919-10930.

Zhou, S., Hu, G., Zhang, Z., & Guan, J. (2008). An empirical study of Chinese language networks. *Physica A,* 387, 3039-3047.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Oxford: Addison-Wesley Press.

Appendix A: (1) Top 100 Words in the Untagged Law Corpus

N	Words	Freq.	%	N	Words	Freq.	%
1	THE	74169	7.523	51	MEANS	1934	0.196
2	OF	46478	4.714	52	WHERE	1914	0.194
3	#	35184	3.569	53	EACH	1868	0.189
4	AND	28336	2.874	54	REQUIRED	1865	0.189
5	TO	26214	2.659	55	STATE	1851	0.188
6	IN	22031	2.235	56	INTERNATIONAL	1814	0.184
7	BE	17869	1.812	57	WATER	1739	0.176
8	A	17350	1.760	58	CERTIFICATE	1644	0.167
9	OR	16544	1.678	59	USED	1631	0.165
10	SHALL	16380	1.661	60	TRANSPORT	1624	0.165
11	FOR	13017	1.320	61	PRESSURE	1623	0.165
12	BY	9470	0.961	62	DATE	1597	0.162
13	WITH	8838	0.896	63	AFTER	1587	0.161
14	IS	7844	0.796	64	OIL	1577	0.160
15	AS	7628	0.774	65	SPACES	1572	0.159
16	THAT	6980	0.708	66	THEIR	1561	0.158
17	ON	6660	0.676	67	SYSTEM	1555	0.158
18	NOT	6060	0.615	68	SECURITY	1513	0.153
19	THIS	5992	0.608	69	SAFETY	1510	0.153
20	WHICH	5799	0.588	70	ADMINISTRATION	1507	0.153
21	ARE	4953	0.502	71	TEST	1484	0.151
22	SHIP	4894	0.496	72	HAS	1460	0.148
23	AT	4464	0.453	73	EQUIPMENT	1457	0.148
24	ANY	4220	0.428	74	PORT	1438	0.146
25	MAY	4085	0.414	75	APPROPRIATE	1424	0.144
26	AN	4025	0.408	76	LESS	1411	0.143
27	SUCH	3813	0.387	77	ARTICLE	1404	0.142
28	CONVENTION	3579	0.363	78	PARTY	1387	0.141
29	FROM	3537	0.359	79	MORE	1355	0.137
30	OTHER	3434	0.348	80	ONE	1325	0.134
31	CARGO	3191	0.324	81	FOLLOWING	1310	0.133
32	THAN	3136	0.318	82	TANK	1282	0.130
33	SHIPS	3113	0.316	83	DANGEROUS	1276	0.129
34	PROVISIONS	2509	0.254	84	MATERIAL	1252	0.127
35	IT	2447	0.248	85	CONTROL	1228	0.125
36	ALL	2444	0.248	86	GOODS	1209	0.123
37	UNDER	2364	0.240	87	FIRE	1207	0.122
38	IF	2337	0.237	88	BEEN	1204	0.122
39	WHEN	2335	0.237	89	SUBSTANCES	1151	0.117
40	PROVIDED	2321	0.235	90	ENSURE	1144	0.116
41	REQUIREMENTS	2208	0.224	91	INCLUDING	1143	0.116
42	HAVE	2150	0.218	92	NO	1137	0.115
43	PARAGRAPH	2138	0.217	93	INFORMATION	1108	0.112
44	ITS	2091	0.212	94	PARTIES	1098	0.111
45	ORGANIZATION	2078	0.211	95	SYSTEMS	1097	0.111
46	SHOULD	2067	0.210	96	WITHIN	1094	0.111
47	INTO	2055	0.208	97	BALLAST	1080	0.110
48	REGULATION	1986	0.201	98	PART	1080	0.110
49	CODE	1964	0.199	99	TYPE	1080	0.110
50	ACCORDANCE	1941	0.197	100	LEAST	1075	0.109

Appendix A: (2) Top 100 Words in the Tagged Law Corpus

N	Words	Freq.	%	N	Words	Freq.	%
1	THE	74060	7.64	51	WHERE	1914	0.20
2	OF	45993	4.74	52	EACH	1868	0.19
3	#	35180	3.63	53	REQUIRED	1865	0.19
4	AND	28232	2.91	54	CODE	1860	0.19
5	TO	26132	2.70	55	STATE	1647	0.17
6	IN	21969	2.27	56	USED	1631	0.17
7	BE	17869	1.84	57	PRESSURE	1586	0.16
8	A	17344	1.79	58	AFTER	1579	0.16
9	OR	16497	1.70	59	TRANSPORT	1565	0.16
10	SHALL	16380	1.69	60	THEIR	1561	0.16
11	FOR	12898	1.33	61	INTERNATIONAL	1552	0.16
12	BY	9465	0.98	62	CERTIFICATE	1519	0.16
13	WITH	8414	0.87	63	SECURITY	1513	0.16
14	IS	7844	0.81	64	DATE	1506	0.16
15	AS	7628	0.79	65	ADMINISTRATION	1505	0.16
16	THAT	6980	0.72	66	SYSTEM	1499	0.15
17	NOT	6056	0.62	67	TEST	1475	0.15
18	THIS	5992	0.62	68	HAS	1460	0.15
19	WHICH	5799	0.60	69	APPROPRIATE	1424	0.15
20	ON	5788	0.60	70	EQUIPMENT	1416	0.15
21	ARE	4953	0.51	71	LESS	1411	0.15
22	SHIP	4790	0.49	72	ARTICLE	1404	0.14
23	ANY	4220	0.44	73	SAFETY	1399	0.14
24	AT	4167	0.43	74	PARTY	1386	0.14
25	MAY	4085	0.42	75	MORE	1355	0.14
26	AN	4025	0.42	76	ONE	1325	0.14
27	SUCH	3813	0.39	77	FOLLOWING	1310	0.14
28	FROM	3533	0.36	78	SPACES	1285	0.13
29	CONVENTION	3449	0.36	79	BEEN	1204	0.12
30	OTHER	3434	0.35	80	TANK	1154	0.12
31	THAN	3136	0.32	81	ENSURE	1144	0.12
32	SHIPS	3032	0.31	82	INCLUDING	1143	0.12
33	CARGO	2878	0.30	83	NO	1136	0.12
34	PROVISIONS	2509	0.26	84	INFORMATION	1101	0.11
35	IT	2447	0.25	85	PARTIES	1096	0.11
36	ALL	2426	0.25	86	WITHIN	1094	0.11
37	UNDER	2357	0.24	87	PART	1079	0.11
38	IF	2337	0.24	88	LEAST	1075	0.11
39	WHEN	2335	0.24	89	SPECIFIED	1073	0.11
40	PROVIDED	2321	0.24	90	APPROVAL	1069	0.11
41	REQUIREMENTS	2208	0.23	91	TYPE	1068	0.11
42	HAVE	2150	0.22	92	SO	1058	0.11
43	PARAGRAPH	2138	0.22	93	SYSTEMS	1054	0.11
44	ITS	2091	0.22	94	AUTHORITY	1051	0.11
45	SHOULD	2067	0.21	95	ANNEX	1043	0.11
46	INTO	2055	0.21	96	SUBSTANCES	1040	0.11
47	ORGANIZATION	2016	0.21	97	MSC	1014	0.10
48	REGULATION	1985	0.20	98	THOSE	1014	0.10
49	ACCORDANCE	1941	0.20	99	THESE	1010	0.10
50	MEANS	1927	0.20	100	USE	1008	0.10

Appendix B: (1) Multi-word Compounds from Wordlist of a Law Corpus

N	Words	Gram
1	INTERNATIONAL CONVENTION FOR THE SAFETY OF LIFE AT SEA	9
2	INTERNATIONAL CONVENTION ON CIVIL LIABILITY FOR OIL CONVENTION	8
3	VESSEL RESTRICTED IN HER ABILITY TO MANOEUVRE	7
4	INTERNATIONAL REGULATIONS FOR PREVENTING COLLISIONS AT SEA	7
5	ELECTRONIC CHART DISPLAY AND INFORMATION SYSTEM	6
6	CERTIFICATE OF PROFICIENCY IN SURVIVAL CRAFT	6
7	A VESSEL CONSTRAINED BY HER DRAUGHT	6
8	INTERNATIONAL CONVENTION ON LOAD LINES	5
9	POLLUTION OF THE MARINE ENVIRONMENT	5
10	INTERNATIONAL OIL POLLUTION PREVENTION CERTIFICATE	5
11	INTERNATIONAL MARITIME DANGEROUS GOODS CODE	5
12	INTERNATIONAL ASSOCIATION OF CLASSIFICATION SOCIETIES	5
13	INTERNATIONAL MEDICAL GUIDE FOR SHIPS	5
14	MEDICAL FIRST AID GUIDE	4
15	INTERNATIONAL CODE OF SIGNALS	4
16	TURN OF THE BILGE	4
17	CLOSEST POINT OF APPROACH	4
18	AUTOMATIC RADAR PLOTTING AIDS	4
19	INSTANTANEOUS RATE OF DISCHARGE	4
20	INTERNATIONAL CHAMBER OF SHIPPING	4
21	DIRECTION OF TRAFFIC FLOW	4
22	FIRE IN THE CARGO	4
23	AUTOMATIC RADAR PLOTTING AID	4
24	FORE AND AFT LINE	4
25	FORE AND AFTER PEAK	4
26	FORWARD OF THE BEAM	4
27	GLOBAL NAVIGATION SATELLITE SYSTEM	4
28	LOADING AND DISCHARGING CARGO	4
29	OFFICER OF THE WATCH	4
30	SIGNALS TO ATTRACT ATTENTION	4
31	SEARCH AND RESCUE	3
32	PREVENTION OF POLLUTION	3
33	INTERNATIONAL MARITIME ORGANIZATION	3
34	NOXIOUS LIQUID SUBSTANCES	3
35	PORT STATE CONTROL	3
36	CERTIFICATE OF FITNESS	3
37	PROCEED TO SEA	3
38	OIL RECORD BOOK	3
39	ANGLE OF REPOSE	3
40	PORT OF CALL	3
41	COUNTRY OF ORIGIN	3
42	CENTRE OF GRAVITY	3
43	INERT GAS SYSTEM	3
44	INERT GAS SYSTEMS	3
45	CRUDE OIL WASHING	3
46	INTERNAL COMBUSTION ENGINES	3
47	SOURCE OF IGNITION	3
48	CERTIFICATE OF COMPETENCY	3
49	EXCLUSIVE ECONOMIC ZONE	3
50	LIQUEFIED GAS TANKER	3

N	Words	Gram
51	LIQUEFIED GAS TANKERS	3
52	PORT OF REGISTRY	3
53	SPECIAL DRAWING RIGHT	3
54	LOADING AND DISCHARGE	3
55	NAME OF SHIP	3
56	PARALLEL OF LATITUDE	3
57	AUXILIARY STEERING GEAR	3
58	GENERAL PURPOSE CONTAINER	3
59	INTERNATIONAL TONNAGE CERTIFICATE	3
60	INTERNAL COMBUSTION ENGINE	3
61	DIGITAL SELECTIVE CALLING	3
62	ELECTRONIC DATA INTERCHANGE	3
63	ELECTRONIC DATA PROCESSING	3
64	FREE SURFACE EFFECT	3
65	MEANS OF PROPULSION	3
66	PORT OF ARRIVAL	3
67	PORT OF DESTINATION	3
68	PORT OF LOADING	3
69	SAFETY EQUIPMENT CERTIFICATE	3
70	TIMBER DECK CARGOES	3
71	AIDS TO NAVIGATION	3
72	FIRE CONTROL PLANS	3
73	STATE OF ORIGIN	3
74	TIMBER DECK CARGO	3
75	FACTOR OF SAFETY	3
76	FORWARD AND AFT	3
77	HEAVY DIESEL OIL	3
78	LOADING AND DISCHARGING	3
79	NOXIOUS LIQUID SUBSTANCE	3
80	NUMBER OF CREW	3
81	TRAFFIC SEPARATION SCHEMES	3
82	ALTERATION OF COURSE	3
83	DEGREE OF SUBDIVISION	3
84	EMERGENCY POWER SUPPLY	3
85	FIRE MAIN SYSTEM	3
86	GEF INTERW PROJECT	3
87	INTERNATIONAL SHORE CONNECTION	3
88	MAXIMUM GROSS WEIGHT	3
89	PARTIALLY ENCLOSED LIFEBOATS	3
90	PRESSURE RELIEF VALVES	3
91	SAFE WORKING LOAD	3
92	SHIFT OF CARGO	3
93	TRAFFIC SEPARATION SCHEME	3
94	TYPE OF CARGO	3
95	TYPE OF SHIP	3
96	ABAFT THE BEAM	3
97	COORDINATED UNIVERSAL TIME	3
98	DANGEROUS GOODS DECLARATION	3
99	DECISION SUPPORT SYSTEM	3
100	FREE SURFACE EFFECTS	3

N	Words	Gram
101	HIGH WATER LEVEL	3
102	NOT UNDER COMMAND	3
103	PORT OF DISCHARGE	3
104	TANK HATCH COVERS	3
105	VESSEL TRAFFIC SERVICES	3
106	ACKNOWLEDGEMENT OF RECEIPT	3
107	ARC OF VISIBILITY	3
108	BILL OF LADING	3
109	CARGO HANDLING EQUIPMENT	3
110	CONSTRUCTIVE TOTAL LOSSES	3
111	CRUDE OIL WASHED	3
112	DET NORSKE VERITAS	3
113	DRY BULK CONTAINER	3
114	ENGINE CONTROL ROOM	3
115	FRESH WATER ALLOWANCE	3
116	FUEL OIL SUPPLY	3
117	IN THE MARKING	3
118	INSHORE TRAFFIC ZONE	3
119	INTERNATIONAL SHORE CONNECTIONS	3
120	LOAD LINE CONVENTION	3
121	MECHANICAL PILOT HOISTS	3
122	MODE OF TRANSPORT	3
123	OPEN TOP CONTAINER	3
124	OPEN TOP CONTAINERS	3
125	PERMANENT BALLAST TANKS	3
126	POINT OF DEPARTURE	3
127	PRESSURE RELIEF VALVE	3
128	ROOT MEAN SQUARE	3
129	SAFE WORKING LOADS	3
130	SEA WATER PUMP	3
131	SPECIAL DRAWING RIGHTS	3
132	WATERTIGHT TRANSVERSE BULKHEADS	3
133	AFTER PEAK BULKHEAD	3
134	AFTER PEAK TANKS	3
135	ARTICLES OF AGREEMENT	3
136	AUTOMATIC CONTROL SYSTEM	3
137	BY THE STERN	3
138	CARGO CARRYING CAPACITY	3
139	CERTIFICATE OF REGISTRY	3
140	CONTROLLABLE PITCH PROPELLERS	3
141	DAILY SERVICE TANK	3
142	DOUBLE BOTTOM CONSTRUCTION	3
143	DRY BULK CARGO	3
144	ELECTRONIC NAVIGATIONAL CHARTS	3
145	ENHANCED GROUP CALL	3
146	ENHANCED GROUP CALLING	3
147	EXTINGUISH A FIRE	3
148	FIRE CONTROL PLAN	3
149	FIRE FIGHTING APPLIANCES	3
150	GENERAL ARRANGEMENT PLANS	3

N	Words	Gram
151	GENERAL PURPOSE CONTAINERS	3
152	GLOBAL POSITIONING SYSTEM	3
153	INNER BOTTOM PLATING	3
154	LAUNCH THE LIFEBOATS	3
155	LENGTH BETWEEN PERPENDICULARS	3
156	LIQUEFIED NATURAL GAS	3
157	LOAD AND DISCHARGE	3
158	LOW SULPHUR FUEL	3
159	LOWER EXPLOSIVE LIMIT	3
160	PERSONAL FLOTATION DEVICES	3
161	QUICK CLOSING VALVE	3
162	REPATRIATION OF SEAMEN	3
163	RESIDUAL FUEL OIL	3
164	RESIDUAL FUEL OILS	3
165	RSLT LIST ITEM	3
166	SHIP HANDLING SIMULATOR	3
167	SOURCE OF IGNITIONS	3
168	SPONTANEOUS IGNITION TEMPERATURE	3
169	THE OPEN SEA	3
170	TIMBER LOAD LINE	3
171	TYPE OF EQUIPMENT	3
172	TYPE OF SHIPS	3
173	UNIT LOAD DEVICES	3
174	VESSEL IN DISTRESS	3
175	VESSEL TRAFFIC SERVICE	3
176	VISIBILITY OF LIGHTS	3
177	VOLUME OF DISPLACEMENT	3
178	WATER STOWAWAYS BROCHURE	3
179	WATER TUBE BOILERS	3
180	WEIGHTS AND MEASURES	3
181	WHO CONSTITUTION EN	3
182	DANGEROUS GOODS	2
183	SUBJECT TO	2
184	BALLAST WATER	2
185	ON BOARD	2
186	COMPLY WITH	2
187	REFER TO	2
188	APPLY TO	2
189	ACCORDING TO	2
190	FITTED WITH	2
191	GROSS TONNAGE	2
192	PRIOR TO	2
193	MACHINERY SPACES	2
194	OIL TANKERS	2
195	RADIOACTIVE MATERIAL	2
196	BULK CARGOES	2
197	DUE TO	2
198	OIL POLLUTION	2
199	AT SEA	2
200	PORT STATE	2

N	Words	Gram	N	Words	Gram
201	PACKING INSTRUCTION	2	251	DAMAGE STABILITY	2
202	MACHINERY SPACE	2	252	POWER SUPPLY	2
203	SURVIVAL CRAFT	2	253	WATERTIGHT DOORS	2
204	OIL TANKER	2	254	FRESH WATER	2
205	NORMAL CONDITIONS	2	255	ANNUAL SURVEYS	2
206	CONSIST OF	2	256	CARGO HOLD	2
207	INERT GAS	2	257	CHEMICAL TANKERS	2
208	EQUIPPED WITH	2	258	SECOND ENGINEER	2
209	STEERING GEAR	2	259	TENSILE STRENGTH	2
210	EXPIRY DATE	2	260	BREATHING APPARATUS	2
211	FIRE DETECTION	2	261	EXPOSED TO	2
212	LIQUEFIED GASES	2	262	ANNUAL SURVEY	2
213	MOISTURE CONTENT	2	263	CARGO HANDLING	2
214	PROPULSION MACHINERY	2	264	INTERNAL COMBUSTION	2
215	RESCUE BOAT	2	265	PORT STATES	2
216	BULK CARGO	2	266	EMERGENCY ESCAPE	2
217	FUEL OIL	2	267	INTERMEDIATE SURVEY	2
218	FIRE ALARM	2	268	VERTICAL ZONE	2
219	RECEPTION FACILITIES	2	269	WEATHER DECK	2
220	BULKHEAD DECK	2	270	CONTROL STATION	2
221	DOUBLE BOTTOM	2	271	NAUTICAL MILES	2
222	PORT FACILITIES	2	272	RHUMB LINE	2
223	BULK CARRIERS	2	273	WATERTIGHT BULKHEADS	2
224	RESCUE BOATS	2	274	CARGO HOLDS	2
225	CRUDE OIL	2	275	RADIO INSTALLATION	2
226	LOAD LINE	2	276	RELATIVE DENSITY	2
227	ON DECK	2	277	AUXILIARY MACHINERY	2
228	FREIGHT CONTAINERS	2	278	FREEBOARD DECK	2
229	CONTROL STATIONS	2	279	RADIO REGULATIONS	2
230	HARMFUL SUBSTANCES	2	280	GENERAL CARGO	2
231	IMDG CODE	2	281	CLEAN BALLAST	2
232	CHIEF ENGINEER	2	282	FIRE CONTROL	2
233	SEGREGATED BALLAST	2	283	FIRE PROTECTION	2
234	PORT OR	2	284	MELTING POINT	2
235	SHELL PLATING	2	285	SERIAL NUMBER	2
236	CHEMICAL TANKER	2	286	ALLOW FOR	2
237	AMBIENT TEMPERATURE	2	287	LUBRICATING OIL	2
238	FIRST AID	2	288	MOULDED DEPTH	2
239	TRANSPORT DOCUMENT	2	289	PACKING INSTRUCTIONS	2
240	FIRE EXTINGUISHERS	2	290	PROOF STRENGTH	2
241	BALLAST TANKS	2	291	REMOTE CONTROL	2
242	EMERGENCY RESPONSE	2	292	WITHOUT PREJUDICE	2
243	FREIGHT CONTAINER	2	293	CARGO UNITS	2
244	LIQUEFIED GAS	2	294	DIRTY BALLAST	2
245	ACCOMMODATION SPACES	2	295	EXISTING SHIPS	2
246	RESTRICTED AREAS	2	296	FIRE DAMPER	2
247	FIRE PUMPS	2	297	FIRE EXTINGUISHING	2
248	SLOP TANKS	2	298	NEAREST LAND	2
249	DIESEL ENGINE	2	299	RECEPTION FACILITY	2
250	TERRITORIAL SEA	2	300	WING TANKS	2

N	Words	Gram	N	Words	Gram
301	ATMOSPHERIC PRESSURE	2	351	OIL TANK	2
302	DEVELOPING COUNTRIES	2	352	OIL TANKS	2
303	HATCH COVERS	2	353	SWAP BODIES	2
304	IN BALLAST	2	354	ADJACENT SPACES	2
305	AIR POLLUTION	2	355	AIR PIPES	2
306	BULK CARRIER	2	356	DEPEND ON	2
307	LOAD LINES	2	357	DOUBLE HULL	2
308	TANK CLEANING	2	358	FIRE PREVENTION	2
309	VERTICAL ZONES	2	359	FLASH POINT	2
310	WATER SPRAY	2	360	FLUE GAS	2
311	MAJOR CONVERSION	2	361	MESS ROOMS	2
312	OILY MIXTURES	2	362	SHIP'S STORES	2
313	SAFETY EQUIPMENT	2	363	SOUNDING PIPES	2
314	STORAGE CAPACITY	2	364	TRANSVERSE BULKHEADS	2
315	AT ANCHOR	2	365	WATER BALLAST	2
316	COLLISION BULKHEAD	2	366	AIR CONDITIONING	2
317	DISTRESS ALERTS	2	367	AMBIENT TEMPERATURES	2
318	FIRE PUMP	2	368	DANGEROUS GOOD	2
319	FREE SURFACE	2	369	DELIVERY NOTE	2
320	MERCHANT SHIPPING	2	370	DISTRESS ALERT	2
321	OIL FILTERING	2	371	FIRE EXTINGUISHER	2
322	ABANDON SHIP	2	372	FORWARD PERPENDICULAR	2
323	EXCLUSIVE USE	2	373	IN TRANSIT	2
324	FORKLIFT POCKETS	2	374	MARITIME LAW	2
325	IOPP CERTIFICATE	2	375	NAVIGATIONAL AIDS	2
326	OILY MIXTURE	2	376	PIPING ARRANGEMENTS	2
327	RADIO INSTALLATIONS	2	377	PRIME MOVER	2
328	RUDDER STOCK	2	378	QUALITY ASSURANCE	2
329	TRANSVERSE BULKHEAD	2	379	RESTRICTED VISIBILITY	2
330	WATERTIGHT DOOR	2	380	SAFETY PRECAUTIONS	2
331	COMBINATION CARRIERS	2	381	CLOSING ARRANGEMENTS	2
332	FOAM SOLUTION	2	382	DIESEL OIL	2
333	OVER ALL	2	383	DISTRESS INCIDENTS	2
334	SLOP TANK	2	384	DRINKING WATER	2
335	TANK DECK	2	385	DRY CARGO	2
336	WATER LEVEL	2	386	EMERGENCY STEERING	2
337	BILGE PUMPING	2	387	FIRE DAMPERS	2
338	BILGE PUMPS	2	388	HAZARDOUS AREAS	2
339	BLANK FLANGE	2	389	INCLINING TEST	2
340	EN ROUTE	2	390	ISOLATING VALVES	2
341	EXEMPTION CERTIFICATE	2	391	NEW SHIP	2
342	FLAMMABLE SOLIDS	2	392	NEW SHIPS	2
343	UNIT LOAD	2	393	PRODUCT CARRIER	2
344	BILGE PUMP	2	394	SET UP	2
345	BILGE WATER	2	395	SMOKE SIGNALS	2
346	CARRYING CAPACITY	2	396	SOUND SIGNALS	2
347	COASTAL TRADING	2	397	TANK TOP	2
348	EMERGENCY GENERATOR	2	398	VHF RADIO	2
349	EXISTING SHIP	2	399	WEATHER CONDITIONS	2
350	MAIN ENGINE	2	400	ENGINE DEPARTMENT	2

N	Words	Gram	N	Words	Gram
401	HOLDING TANK	2	451	MASTHEAD LIGHT	2
402	INNER BOTTOM	2	452	MASTHEAD LIGHTS	2
403	ISM CODE	2	453	OFFSHORE INSTALLATION	2
404	ROUTEING SYSTEMS	2	454	OPERATING PROCEDURES	2
405	SAFE SPEED	2	455	PILOT LADDER	2
406	SALT WATER	2	456	PORT AUTHORITIES	2
407	SPONTANEOUS COMBUSTION	2	457	RADIO BEACONS	2
408	WEATHERTIGHT DOORS	2	458	RADIOACTIVE MATERIALS	2
409	ABLE SEAMEN	2	459	RESPONSIBLE OFFICER	2
410	ACCOMMODATION LADDER	2	460	RO-RO SHIPS	2
411	ADVERSE WEATHER	2	461	SAFE PASSAGE	2
412	ALL-ROUND LIGHTS	2	462	SHIPPING DOCUMENTS	2
413	BALLAST TANK	2	463	STOWAGE FACTOR	2
414	CORNER POSTS	2	464	SUCTION PIPE	2
415	DANGEROUS CARGO	2	465	TANK CONTAINERS	2
416	DECK DEPARTMENT	2	466	BASE LINE	2
417	DIESEL ENGINES	2	467	BLANK FLANGES	2
418	ELECTRIC LIGHTING	2	468	DISCHARGE PIPING	2
419	FREE SURFACES	2	469	DISTRESS SIGNALS	2
420	FUEL SUPPLY	2	470	DRY BULK	2
421	GALVANIC ACTION	2	471	EXPLOSION PROOF	2
422	HARMFUL SUBSTANCE	2	472	FINAL DESTINATION	2
423	HIGH PRESSURE	2	473	FUEL OILS	2
424	HYDRAULIC FLUID	2	474	HAZARDOUS AREA	2
425	INTERMEDIATE SURVEYS	2	475	INSULATION SYSTEM	2
426	INTERNAL WATERS	2	476	LIFEJACKET LIGHT	2
427	MEDICAL ADVICE	2	477	MAKING WAY	2
428	MESS ROOM	2	478	MARINE CASUALTIES	2
429	METACENTRIC HEIGHT	2	479	MOULDED BREADTH	2
430	NARROW CHANNEL	2	480	MUCOUS MEMBRANES	2
431	NUCLEAR SHIP	2	481	PERFORMANCE INDICATORS	2
432	PASSENGER LIST	2	482	PIPE FITTINGS	2
433	PRIMARY BARRIER	2	483	PIPING ARRANGEMENT	2
434	REFRIGERATED CARGO	2	484	PROPULSION SYSTEM	2
435	SLUDGE TANK	2	485	PUMP ROOMS	2
436	STEEL TUBES	2	486	QUALITY CONTROL	2
437	TRANSPORT DOCUMENTS	2	487	RATED CAPACITY	2
438	VHF CHANNEL	2	488	ROAD VEHICLES	2
439	AIR COMPRESSOR	2	489	SEA CHEST	2
440	AUTOMATIC IDENTIFICATION	2	490	SEA CONDITIONS	2
441	CLASSIFICATION SOCIETIES	2	491	SPECIFIC GRAVITY	2
442	CONTRARY TO	2	492	STOWAGE PLAN	2
443	CREW LIST	2	493	SUMMER FREEBOARD	2
444	DEAD SHIP	2	494	SUPPLY VESSELS	2
445	DECK CARGO	2	495	TAKE IN	2
446	DEPEND UPON	2	496	VEGETABLE OIL	2
447	EFFECTIVE DATE	2	497	WATER FOG	2
448	FLOW CHART	2	498	WATERTIGHT BULKHEAD	2
449	INSURANCE CERTIFICATE	2	499	WING TANK	2
450	LIQUID LEVEL	2	500	ABOUT SHIPS	2

N	Words	Gram	N	Words	Gram
501	ACCOMMODATION SPACE	2	551	SOUND SIGNAL	2
502	ANCHOR WINDLASS	2	552	SOUND SIGNALLING	2
503	BACK OF	2	553	SPARE PARTS	2
504	BOILER ROOM	2	554	TOPIC ID	2
505	BUOYANT APPARATUS	2	555	TOWING VESSEL	2
506	CENTRE TANK	2	556	TRAFFIC SEPARATION	2
507	CHAIN LOCKER	2	557	TRIPPING LINE	2
508	COMBINATION CARRIER	2	558	ULTIMATE STRENGTH	2
509	CONTAINER SHIPS	2	559	UNIT LOADS	2
510	CONTINGENCY PLAN	2	560	UPPER DECK	2
511	CONTROLLABLE PITCH	2	561	VENT PIPING	2
512	CUBIC FEET	2	562	WATER SPRAYING	2
513	DATA PLATE	2	563	WIND FORCE	2
514	DECK MACHINERY	2	564	AIR COMPRESSORS	2
515	DYNAMIC POSITIONING	2	565	BLIND SECTORS	2
516	ELECTRIC MOTOR	2	566	BOTTOM PLATING	2
517	ELECTRIC MOTORS	2	567	BROUGHT UP	2
518	ENGINE ROOM	2	568	CALORIFIC VALUE	2
519	EXHAUST GAS	2	569	CARGO BLOCK	2
520	EXPOSED DECK	2	570	CARGO HATCH	2
521	FISHING GEAR	2	571	COMPASS BEARING	2
522	FISHING VESSELS	2	572	COMPASS ERROR	2
523	FREEING PORTS	2	573	COMPOSITE UNIT	2
524	HAND FLARES	2	574	CONNING POSITION	2
525	HOT WORK	2	575	COOLING SYSTEM	2
526	HYPOTHETICAL OUTFLOW	2	576	COOLING SYSTEMS	2
527	ICE CONDITIONS	2	577	CROSS MEMBERS	2
528	INLAND PORTS	2	578	CRUISE SHIP	2
529	LIFEJACKET LIGHTS	2	579	DISCHARGE PIPE	2
530	LOADING PORTS	2	580	DISCHARGE PORTS	2
531	LONG TON	2	581	DOUBLE BOTTOMS	2
532	LOW WATER	2	582	DRILLING RIG	2
533	MAIN ENGINES	2	583	DUCT KEELS	2
534	MEDICAL GUIDE	2	584	ENGINE ROOMS	2
535	NUCLEAR SHIPS	2	585	FEED WATER	2
536	ORE CARRIERS	2	586	FOAM EXTINGUISHERS	2
537	ORGANIC LIQUIDS	2	587	FOOD RATION	2
538	PORT SIDE	2	588	FOREPEAK TANK	2
539	POSITION FIXING	2	589	GALLEY RANGES	2
540	POWER SUPPLIES	2	590	GAS FREEING	2
541	PRO RATA	2	591	HAND FLARE	2
542	REFRIGERATING MACHINES	2	592	HEATING COILS	2
543	RELATIVE DENSITIES	2	593	IGNITION TEMPERATURE	2
544	ROAD VEHICLE	2	594	INLAND WATERWAYS	2
545	SAFETY VALVES	2	595	KINEMATIC VISCOSITY	2
546	SET IN	2	596	LIFEBOAT EQUIPMENT	2
547	SEVERE DAMAGE	2	597	LIQUID FUELLED	2
548	SHIP'S AGENT	2	598	LUBRICATION OIL	2
549	SHIP'S OFFICER	2	599	MAIN INLETS	2
550	SHORT BLASTS	2	600	NET TONNAGES	2

N	Words	Gram	N	Words	Gram
601	OFFSHORE INSTALLATIONS	2	651	AUTOMATIC ACQUISITION	2
602	OILY WATER	2	652	AUTOMATIC PILOT	2
603	OWNER CODE	2	653	BALLAST PUMPS	2
604	PAD EYES	2	654	BALLAST WATERS	2
605	PARACHUTE FLARES	2	655	BAROMETRIC PRESSURE	2
606	PASSENGER LISTS	2	656	BEAUFORT SCALE	2
607	PORT SERVICES	2	657	BENDING STRESS	2
608	POWER FAILURE	2	658	BETWEEN DECK	2
609	PROLONGED BLASTS	2	659	BILGE WATERS	2
610	PROPULSION PLANT	2	660	BLIND SECTOR	2
611	PUMP ROOM	2	661	BOILER ROOMS	2
612	RADAR PLOTTING	2	662	BOW DOOR	2
613	RADIO ROOM	2	663	BOW DOORS	2
614	RADIO STATION	2	664	BREAK BULK	2
615	RESTRICTED AREA	2	665	BUOYANT LIFELINE	2
616	ROUTEING SYSTEM	2	666	BUOYANT LIFELINES	2
617	RUDDER ANGLE	2	667	CARGO EQUIPMENT	2
618	SHIP OPERATOR	2	668	CELESTIAL BODIES	2
619	SHIP OWNER	2	669	CLEAN CARGO	2
620	SHIP'S LENGTH	2	670	COMBUSTIBLE LIQUIDS	2
621	SHIP'S OFFICERS	2	671	CONSEQUENTIAL LOSS	2
622	SHORE CONNECTION	2	672	CONTAINER SHIP	2
623	SHUT DOWN	2	673	CONTINGENCY PLANNING	2
624	SIDE PLATING	2	674	COOLING MEDIA	2
625	SIDE SCUTTLES	2	675	CREW'S QUARTERS	2
626	SISTER SHIP	2	676	CRUDE OILS	2
627	SMOKE SIGNAL	2	677	DATA ID	2
628	SOLVENT EXTRACTIONS	2	678	DECK HEIGHT	2
629	STARBOARD SIDE	2	679	DELIVERY NOTES	2
630	STATIC ELECTRICITY	2	680	DETECTION RANGES	2
631	STORM WARNING	2	681	DIRECT ROUTES	2
632	TERRITORIAL SEAS	2	682	DOC ID	2
633	THERMAL CONTAINER	2	683	DREDGED MATERIAL	2
634	THERMAL CONTAINERS	2	684	ELECTRIC CABLES	2
635	TIDE TABLES	2	685	ELECTRONIC NAVIGATION	2
636	TIME ZONES	2	686	END ON	2
637	TRAFFIC DENSITY	2	687	ENVIRONMENTAL LOADS	2
638	TRAFFIC FLOW	2	688	ENVIRONMENTAL POLLUTION	2
639	TRAFFIC LANE	2	689	EVEN KEEL	2
640	ULLAGE SPACE	2	690	EXTREME BREADTH	2
641	VENTILATED CONTAINER	2	691	FINANCIAL RESPONSIBILITY	2
642	WITHOUT RECOURSE	2	692	FIRE SIGNALS	2
643	YIELD POINT	2	693	FIRST WATCH	2
644	AD HOC	2	694	FLAME ARRESTERS	2
645	AFTER PERPENDICULARS	2	695	FLAMMABLE SOLID	2
646	AIR CONDITIONS	2	696	FLASHING LIGHT	2
647	ALL IN	2	697	FLOOR PLATING	2
648	ALTERNATING CURRENT	2	698	FORECASTLE HEAD	2
649	ANCHOR CHAIN	2	699	FREEBOARD MARKS	2
650	ASSISTANT ENGINEER	2	700	FULL RUDDER	2

N	Words	Gram	N	Words	Gram
701	GALLEY RANGE	2	751	PRODUCT CARRIERS	2
702	GENERAL PERMITS	2	752	RADIO BEACON	2
703	GROSS WEIGHT	2	753	RADIO STATIONS	2
704	HATCH COVER	2	754	REFRIGERATING MACHINERY	2
705	HOLDING ON	2	755	REPORTING FORM	2
706	HOMOGENEOUS CARGO	2	756	SAFETY VALVE	2
707	HORIZONTAL PLANE	2	757	SEA STATE	2
708	INSHORE TRAFFIC	2	758	SEA-GOING VESSEL	2
709	INSULATED TANK	2	759	SHAFT ALLEYS	2
710	INSULATION MATERIAL	2	760	SHALLOW WATER	2
711	INSULATION SYSTEMS	2	761	SHELL CAPACITY	2
712	INTERNAT RESPONSE	2	762	SHIPPING MARKS	2
713	INTRINSICALLY SAFE	2	763	SHORT BLAST	2
714	IOC DOC	2	764	SHORT SEA	2
715	LATENT HEAT	2	765	SIGNAL LIGHTS	2
716	LENGTH OVERALL	2	766	SLUDGE TANKS	2
717	LEVEL ICE	2	767	SPEC PUB	2
718	LINE-THROWING APPLIANCE	2	768	STABLE EQUILIBRIUM	2
719	LINE-THROWING APPLIANCES	2	769	STEAM TURBINES	2
720	LIQUID FUEL	2	770	STEEL PLATE	2
721	LNG CARRIER	2	771	STEERING GEARS	2
722	LOWER HOLD	2	772	STEERING WHEEL	2
723	MAIN DECK	2	773	STERN DOORS	2
724	MAKING UP	2	774	STERN TUBE	2
725	MANIFOLD VALVE	2	775	TANK CONTAINER	2
726	MANUAL STEERING	2	776	TANK TOPS	2
727	MARITIME ADMINISTRATION	2	777	TELEPHONE NUMBER	2
728	MELTING POINTS	2	778	TIDAL RANGE	2
729	MEMBRANE TANKS	2	779	TIDAL STREAM	2
730	METRIC TONS	2	780	TRAFFIC MANAGEMENT	2
731	MIDSHIP SECTION	2	781	TWIST LOCKS	2
732	MINERAL OIL	2	782	UNDER WAY	2
733	NAVIGABLE WATERS	2	783	VACUUM VALVE	2
734	NAVIGATION LIGHTS	2	784	VACUUM VALVES	2
735	NAVIGATION MARK	2	785	VENT PIPE	2
736	NAVIGATIONAL AID	2	786	VENT PIPES	2
737	NOTIFY PARTIES	2	787	VENTILATED CONTAINERS	2
738	ON SHORE	2	788	WATERTIGHT COMPARTMENT	2
739	OPEN WATER	2	789	WEATHER ROUTEING	2
740	ORGANIC MATERIALS	2	790	WIRE ROPE	2
741	OUTBOARD MOTOR	2	791	ADJACENT SPACE	2
742	OVERTAKING VESSEL	2	792	ADVANCE PAYMENT	2
743	OXIDIZING MATERIALS	2	793	AFT DECK	2
744	PAID VACATIONS	2	794	AFTER PART	2
745	PARACHUTE FLARE	2	795	AFTER PEAK	2
746	PILOT VESSEL	2	796	AFTER PERPENDICULAR	2
747	PORT AREAS	2	797	AIR WAYBILL	2
748	PORT CHARGES	2	798	ALL WATER	2
749	PORTABLE EXTINGUISHERS	2	799	ALL WATERS	2
750	POTABLE WATER	2	800	ALTER COURSE	2

N	Words	Gram	N	Words	Gram
801	ANCHOR WATCH	2	851	DRAUGHT MARKS	2
802	AUXILIARY ENGINES	2	852	DREDGE NET	2
803	BALLAST PUMPING	2	853	DRILLING RIGS	2
804	BANK GUARANTEE	2	854	DRY DOCK	2
805	BAR KEEL	2	855	DUCT KEEL	2
806	BARGE CARRIER	2	856	ECHO SOUNDING	2
807	BAROMETRIC TENDENCY	2	857	EDIBLE OIL	2
808	BASIC RATE	2	858	EFFECTIVE POWER	2
809	BAY PLAN	2	859	ELECTRIC CIRCUITS	2
810	BENZENE RINGS	2	860	ELECTRIC LIGHT	2
811	BERTHING CAPACITY	2	861	ELECTRIC PROPULSION	2
812	BLOCK STOWAGE	2	862	ELECTRICAL CIRCUIT	2
813	BOARDING LADDER	2	863	ELECTRICAL CIRCUITS	2
814	BOTTOM PLATE	2	864	EMERGENCY GENERATORS	2
815	BREATHING APPARATUSES	2	865	EMERGENCY MESSAGES	2
816	BUTT JOINT	2	866	ESCAPE SLIDE	2
817	BWTT JUNE	2	867	EXPANSION JOINT	2
818	CARGO GEAR	2	868	EXPANSION JOINTS	2
819	CARGO MANIFOLD	2	869	EXPIRY DATES	2
820	CARGO SEGREGATION	2	870	EXPLOSIVE LIMITS	2
821	CATHODIC PROTECTION	2	871	FALSE ECHOES	2
822	CHART DATA	2	872	FEEDER VESSEL	2
823	CLASSIFICATION SOCIETY	2	873	FILLING PIPES	2
824	CLOSE QUARTERS	2	874	FIRE ALARMS	2
825	COMPARMENT AT	2	875	FIRE APPLIANCES	2
826	COMPASS ERRORS	2	876	FIRE DAMAGE	2
827	CONTAINER PREFIX	2	877	FIRE POINT	2
828	CONTRACT CARRIERS	2	878	FIRE SIGNAL	2
829	CONTROL POINTS	2	879	FIRST LINE	2
830	COOLING MEDIUM	2	880	FISHING VESSEL	2
831	CORNER CASTING	2	881	FITTING OUT	2
832	CORRECTION FACTOR	2	882	FLAME SCREENS	2
833	CORRUGATED BULKHEAD	2	883	FLAMMABLE RANGE	2
834	CORRUGATED BULKHEADS	2	884	FLASH TUBE	2
835	CRUISE SHIPS	2	885	FLOOR PLATES	2
836	CUSTOMS CLEARANCE	2	886	FOAM EXTINGUISHER	2
837	DANGEROUS CARGOES	2	887	FOG SIGNAL	2
838	DATA PLATES	2	888	FOG SIGNALS	2
839	DECK LINE	2	889	FOOD RATIONS	2
840	DECK MAINTENANCE	2	890	FOREIGN-GOING SHIP	2
841	DEEP TANK	2	891	FOREPEAK TANKS	2
842	DEEP TANKS	2	892	FREE CIRCULATION	2
843	DELIVERY LINES	2	893	FREEZING POINT	2
844	DIRECT CURRENT	2	894	FUEL CONSUMPTION	2
845	DIRECT ROUTE	2	895	FUEL FILTERS	2
846	DISCHARGE PIPES	2	896	GAS WELDING	2
847	DISCHARGE PORT	2	897	GENERAL ARRANGEMENT	2
848	DISTILLATION RANGE	2	898	GENERAL CARGOES	2
849	DISTRESS ALERTING	2	899	GOOD FAITH	2
850	DISTRESS CALL	2	900	GRACE PERIOD	2

N	Words	Gram	N	Words	Gram
901	HATCH COAMINGS	2	951	ON BOARDING	2
902	HOLD BACK	2	952	OPEN GAUGING	2
903	HOLDING TANKS	2	953	ORGANIC MATERIAL	2
904	HORIZONTAL POSITION	2	954	OWNER CODES	2
905	HYDROGRAPHIC SURVEYING	2	955	OXIDIZING AGENTS	2
906	HYPOTHETICAL OUTFLOWS	2	956	PER SE	2
907	ICE SHELF	2	957	PERFORMANCE INDICATOR	2
908	INERT GASES	2	958	PERFORMANCE MEASUREMENT	2
909	INSULATED TANKS	2	959	PILOT LADDERS	2
910	INTERNAL MOVEMENTS	2	960	PILOT STATION	2
911	INTERNAL WATER	2	961	PILOT VESSELS	2
912	INTERNATIONAL CARRIAGE	2	962	PORT CONTROL	2
913	LEASING COMPANIES	2	963	PORT INDUSTRIES	2
914	LEE SIDE	2	964	PORT SERVICE	2
915	LIFEBOAT ENGINE	2	965	PORTABLE EXTINGUISHER	2
916	LIFESAVING APPLIANCES	2	966	PORTABLE RADIO	2
917	LIGHT SIGNAL	2	967	POWDER EXTINGUISHERS	2
918	LIGHT SIGNALS	2	968	PRIVATE CARRIERS	2
919	LOAD DRAUGHT	2	969	PROLONGED BLAST	2
920	LOAD FACTORS	2	970	PROOF LOAD	2
921	LONG BLAST	2	971	PROPELLER SHAFTING	2
922	LOWER HOLDS	2	972	PROPULSION SYSTEMS	2
923	LUMP SUM	2	973	PUT OUT	2
924	MAGNETIC COMPASS	2	974	QUARTERING SEAS	2
925	MAGNETIC COMPASSES	2	975	RADAR REFLECTORS	2
926	MAIN INLET	2	976	RADIO OFFICER	2
927	MAIN TURBINE	2	977	RADIO REGULATION	2
928	MAIN TURBINES	2	978	RADIO ROOMS	2
929	MAN ROPES	2	979	RADIO SURVEYS	2
930	MARINE INSURANCE	2	980	RAW MATERIALS	2
931	MARINE RESEARCH	2	981	REFRIGERATED CARGOES	2
932	MARINE TERMINALS	2	982	REFRIGERATING MACHINE	2
933	MARITIME LIENS	2	983	REMOTE CONTROLLED	2
934	MAXIMUM SPEED	2	984	REMOTE CONTROLS	2
935	MEAN DRAUGHT	2	985	RENDER ASSISTANCE	2
936	MISSION STATEMENT	2	986	RESPONSES-INVASIVESSPECIES EN	2
937	MIXED LOAD	2	987	RESPONSIBLE OFFICERS	2
938	MIXED LOADS	2	988	RIGHTING ARM	2
939	MTDSG NO	2	989	RUBBING STRIPS	2
940	MULTIMODAL TRANSPORT	2	990	RUDDER ACTUATOR	2
941	NARROW CHANNELS	2	991	RUDDER ACTUATORS	2
942	NAUTICAL ALMANAC	2	992	SAFETY NET	2
943	NAUTICAL MILE	2	993	SEA CONNECTION	2
944	NAVIGATION MARKS	2	994	SEA GOING	2
945	NAVIGATIONAL WARNING	2	995	SECURING CARGO	2
946	NOISE REDUCTION	2	996	SECURING CARGOES	2
947	OBO SHIPS	2	997	SEPARATION LINE	2
948	ODD NUMBER	2	998	SEPARATION ZONE	2
949	OIL MIST	2	999	SERIAL NUMBERS	2
950	OIL SEPARATORS	2	1000	SETTLING TANKS	2

N	Words	Gram	N	Words	Gram
1001	SHAFT TUNNEL	2	1051	WEB FRAMES	2
1002	SHIP OPERATORS	2	1052	WELDING EQUIPMENT	2
1003	SHIP OWNERS	2	1053	WIND SPEED	2
1004	SHIP'S MANIFEST	2	1054	WORKING DAYS	2
1005	SHORT STAY	2			
1006	SISTER SHIPS	2			
1007	SKELETAL CHASSIS	2			
1008	SLACK TANKS	2			
1009	SMOKE HELMET	2			
1010	SOOT BLOWERS	2			
1011	SOUNDING PIPE	2			
1012	SPEED REDUCTION	2			
1013	STANDARD COMPASS	2			
1014	STANDARD DEVIATION	2			
1015	STAND-ON VESSEL	2			
1016	STEAM DRUM	2			
1017	STEEL PLATES	2			
1018	STEEL TUBE	2			
1019	STERN TUBES	2			
1020	STIFF SHIP	2			
1021	STILL WATER	2			
1022	STOPPING DISTANCE	2			
1023	STOWAGE PLANS	2			
1024	STOWAGE SPACE	2			
1025	STRETCH WRAP	2			
1026	SUCTION PIPING	2			
1027	SUPPLY CHAIN	2			
1028	SUPPLY VESSEL	2			
1029	SWAP BODY	2			
1030	SWASH BULKHEADS	2			
1031	TOWING LIGHT	2			
1032	TOWING WINCHES	2			
1033	TRANSVERSE FRAMING	2			
1034	TRUE COURSE	2			
1035	TRUE NORTH	2			
1036	TRUE WIND	2			
1037	TURN TO	2			
1038	UNDER-KEEL CLEARANCE	2			
1039	UPON ARRIVAL	2			
1040	VEGETABLE OILS	2			
1041	VERTICAL ZONING	2			
1042	VISUAL AIDS	2			
1043	WARNING SIGNAL	2			
1044	WARNING SIGNALS	2			
1045	WATCH OFFICERS	2			
1046	WATER LEVELS	2			
1047	WATERTIGHT COMPARTMENTS	2			
1048	WAVE HEIGHT	2			
1049	WEATHER CHARTS	2			
1050	WEATHER DECKS	2			

Appendix B: (2) Multi-word Compounds of Keywords and Linked Keywords

N	Keywords	Freq.	%	RC Freq.	RC %	Keyness	P-Value
1	DANGEROUS_GOODS	922	0.10	97		1998.54	0.000
2	BALLAST_WATER	686	0.07	189		1116.47	0.000
3	ON_BOARD	718	0.07	423	0.01	746.97	0.000
4	RADIOACTIVE_MATERIAL	247	0.03	5		645.96	0.000
5	GROSS_TONNAGE	335	0.03	79		579.97	0.000
6	FITTED_WITH	334	0.03	83		566.78	0.000
7	MACHINERY_SPACES	208	0.02	4		545.50	0.000
8	BULK_CARGOES	177	0.02	14		403.84	0.000
9	PACKING_INSTRUCTION	139	0.01	0		389.54	0.000
10	OIL_TANKERS	198	0.02	60		308.97	0.000
11	SURVIVAL_CRAFT	125	0.01	6		304.98	0.000
12	MACHINERY_SPACE	120	0.01	4		303.22	0.000
13	NORMAL_CONDITIONS	111	0.01	2		292.10	0.000
14	LIQUEFIED_GASES	92		0		257.82	0.000
15	STEERING_GEAR	99	0.01	5		240.16	0.000
16	INERT_GAS	94		3		238.36	0.000
17	BULKHEAD_DECK	79		0		221.39	0.000
18	EXPIRY_DATE	84		2		217.54	0.000
19	PROPULSION_MACHINERY	80		1		213.98	0.000
20	MOISTURE_CONTENT	83		3		208.27	0.000
21	CONTROL_STATIONS	82		3		205.54	0.000
22	FIRE_DETECTION	82		5		194.36	0.000
23	RESCUE_BOATS	76		4		183.48	0.000
24	FREIGHT_CONTAINERS	70		3		172.84	0.000
25	FIRE_ALARM	71		4		170.00	0.000
26	ACCOMMODATION_SPACES	63		1		166.82	0.000
27	PORT_STATE	148	0.02	85		157.07	0.000
28	LOAD_LINE	66		4		156.56	0.000
29	RESCUE_BOAT	77		13		148.81	0.000
30	IMDG_CODE	76		14		143.10	0.000
31	FREIGHT_CONTAINER	55		2		137.94	0.000
32	INTERNATIONAL_CONVENTION_ ON_LOAD_LINES	51		1		133.61	0.000
33	SHELL_PLATING	53		2		132.48	0.000
34	DOUBLE_BOTTOM	71		14		130.83	0.000
35	HARMFUL_SUBSTANCES	61		7		129.82	0.000
36	FIRE_EXTINGUISHERS	46		0		128.91	0.000
37	NOXIOUS_LIQUID_SUBSTANCES	50		2		124.30	0.000
38	ON_DECK	69		15		123.02	0.000
39	WATERTIGHT_DOORS	47		1		122.56	0.000
40	RECEPTION_FACILITIES	72		18		121.88	0.000
41	BREATHING_APPARATUS	43		0		120.50	0.000
42	SEGREGATED_BALLAST	52		4		119.17	0.000
43	OIL_TANKER	116	0.01	70		118.31	0.000
44	FIRST_AID	47		2		116.13	0.000
45	BULK_CARGO	82		31		114.55	0.000
46	AMBIENT_TEMPERATURE	46		2		113.41	0.000
47	TRANSPORT_DOCUMENT	48		4		108.57	0.000
48	FIRE_PUMPS	40		1		103.26	0.000
49	SLOP_TANKS	45		4		100.66	0.000
50	RESTRICTED_AREAS	39		1		100.51	0.000

N	Keywords	Freq.	%	RC Freq.	RC %	Keyness	P-Value
51	BALLAST_TANKS	48		6		100.23	0.000
52	CERTIFICATE_OF_FITNESS	35		0		98.08	0.000
53	VERTICAL_ZONE	32		0		89.68	0.000
54	OIL_POLLUTION	139	0.01	131		89.56	0.000
55	ANNUAL_SURVEYS	31		0		86.87	0.000
56	WATERTIGHT_BULKHEADS	34		1		86.77	0.000
57	TENSILE_STRENGTH	30		0		84.07	0.000
58	ANGLE_OF_REPOSE	29		0		81.27	0.000
59	FREEBOARD_DECK	28		0		78.47	0.000
60	DAMAGE_STABILITY	37		5		75.85	0.000
61	RADIO_INSTALLATION	27		0		75.66	0.000
62	MOULDED_DEPTH	27		0		75.66	0.000
63	INTERMEDIATE_SURVEY	27		0		75.66	0.000
64	WEATHER_DECK	29		1		73.07	0.000
65	RELATIVE_DENSITY	26		0		72.86	0.000
66	NEAREST_LAND	26		0		72.86	0.000
67	LIQUEFIED_GAS	38		7		71.55	0.000
68	CARGO_UNITS	28		1		70.33	0.000
69	PROCEED_TO_SEA	30		2		70.24	0.000
70	SERIAL_NUMBER	25		0		70.06	0.000
71	PACKING_INSTRUCTIONS	25		0		70.06	0.000
72	EMERGENCY_ESCAPE	27		1		67.60	0.000
73	PROOF_STRENGTH	24		0		67.26	0.000
74	FUEL_OIL	84		68		64.83	0.000
75	MELTING_POINT	23		0		64.45	0.000
76	CARGO_HOLD	36		8		63.69	0.000
77	OIL_RECORD_BOOK	32		5		63.20	0.000
78	RADIO_REGULATIONS	32		5		63.20	0.000
79	PORT_FACILITIES	75		56		63.01	0.000
80	CLEAN_BALLAST	22		0		61.65	0.000
81	SPECIAL_DRAWING_RIGHT	22		0		61.65	0.000
82	DIRTY_BALLAST	22		0		61.65	0.000
83	COUNTRY_OF_ORIGIN	37		10		60.69	0.000
84	RHUMB_LINE	26		2		59.58	0.000
85	LUBRICATING_OIL	29		4		59.16	0.000
86	CONTROL_STATION	31		6		57.47	0.000
87	FIRE_PUMP	23		1		56.71	0.000
88	ANNUAL_SURVEY	28		4		56.62	0.000
89	CRUDE_OIL_WASHING	20		0		56.05	0.000
90	INERT_GAS_SYSTEMS	20		0		56.05	0.000
91	FIRE_DAMPER	20		0		56.05	0.000
92	AUXILIARY_MACHINERY	30		6		55.02	0.000
93	CHIEF_ENGINEER	54		33		54.50	0.000
94	INTERNAL_COMBUSTION_ENGINES	24		2		54.29	0.000
95	CARGO_HOLDS	32		8		54.17	0.000
96	RECEPTION_FACILITY	22		1		53.99	0.000
97	RADIO_INSTALLATIONS	19		0		53.25	0.000
98	COMBINATION_CARRIERS	19		0		53.25	0.000
99	SECOND_ENGINEER	30		7		52.14	0.000
100	POWER_SUPPLY	34		11		51.45	0.000

N	Keywords	Freq.	%	RC Freq.	RC %	Keyness	P-Value
101	OILY_MIXTURE	18		0		50.44	0.000
102	VERTICAL_ZONES	18		0		50.44	0.000
103	MAJOR_CONVERSION	22		2		49.02	0.000
104	WING_TANKS	22		2		49.02	0.000
105	ATMOSPHERIC_PRESSURE	20		1		48.57	0.000
106	FIRE_EXTINGUISHING	20		1		48.57	0.000
107	COLLISION_BULKHEAD	20		1		48.57	0.000
108	INTERNAL_COMBUSTION	17		0		47.64	0.000
109	TRANSVERSE_BULKHEAD	17		0		47.64	0.000
110	CENTRE_OF_GRAVITY	23		3		47.55	0.000
111	FIRE_PROTECTION	28		7		47.40	0.000
112	MEDICAL_FIRST_AID_GUIDE	19		1		45.87	0.000
113	SOURCE_OF_IGNITION	16		0		44.84	0.000
114	FORKLIFT_POCKETS	16		0		44.84	0.000
115	OIL_FILTERING	16		0		44.84	0.000
116	PORT_OR	47		31		44.42	0.000
117	WATER_SPRAY	18		1		43.17	0.000
118	LOAD_LINES	18		1		43.17	0.000
119	DISTRESS_ALERTS	18		1		43.17	0.000
120	IOPP_CERTIFICATE	15		0		42.04	0.000
121	EXEMPTION_CERTIFICATE	15		0		42.04	0.000
122	NAME_OF_SHIP	15		0		42.04	0.000
123	WATERTIGHT_DOOR	15		0		42.04	0.000
124	SLOP_TANK	15		0		42.04	0.000
125	INERT_GAS_SYSTEM	22		4		41.59	0.000
126	OILY_MIXTURES	17		1		40.48	0.000
127	LIQUEFIED_GAS_TANKERS	14		0		39.23	0.000
128	WATER_BALLAST	14		0		39.23	0.000
129	LIQUEFIED_GAS_TANKER	14		0		39.23	0.000
130	BLANK_FLANGE	14		0		39.23	0.000
131	FOAM_SOLUTION	14		0		39.23	0.000
132	SWAP_BODIES	14		0		39.23	0.000
133	TANK_DECK	14		0		39.23	0.000
134	BILGE_PUMPING	14		0		39.23	0.000
135	ADJACENT_SPACES	16		1		37.80	0.000
136	RUDDER_STOCK	16		1		37.80	0.000
137	CHEMICAL_TANKER	47		37		37.38	0.000
138	DANGEROUS_GOOD	13		0		36.43	0.000
139	UNIT_LOAD	13		0		36.43	0.000
140	AIR_PIPES	13		0		36.43	0.000
141	FRESH_WATER	34		20		35.40	0.000
142	FLAMMABLE_SOLIDS	15		1		35.12	0.000
143	PREVENTION_OF_POLLUTION	50		44		35.07	0.000
144	FREE_SURFACE	18		3		34.91	0.000
145	SHIP'S_STORES	12		0		33.63	0.000
146	FORWARD_PERPENDICULAR	12		0		33.63	0.000
147	CERTIFICATE_OF_COMPETENCY	14		1		32.45	0.000
148	REMOTE_CONTROL	23		9		31.52	0.000
149	FIRE_DAMPERS	11		0		30.83	0.000
150	DELIVERY_NOTE	11		0		30.83	0.000

N	Keywords	Freq.	%	RC Freq.	RC %	Keyness	P-Value
151	AUXILIARY_STEERING_GEAR	11		0		30.83	0.000
152	PIPING_ARRANGEMENTS	11		0		30.83	0.000
153	SOUNDING_PIPES	11		0		30.83	0.000
154	DISTRESS_ALERT	11		0		30.83	0.000
155	OIL_TANK	13		1		29.79	0.000
156	FLUE_GAS	13		1		29.79	0.000
157	BILGE_PUMP	13		1		29.79	0.000
158	EMERGENCY_RESPONSE	39		32		29.65	0.000
159	AMBIENT_TEMPERATURES	10		0		28.02	0.000
160	HAZARDOUS_AREAS	10		0		28.02	0.000
161	PARALLEL_OF_LATITUDE	10		0		28.02	0.000
162	SMOKE_SIGNALS	10		0		28.02	0.000
163	FIRE_CONTROL	20		8		27.07	0.000
164	ABANDON_SHIP	15		4		24.74	0.000
165	GENERAL_PURPOSE_CONTAINER	11		1		24.51	0.000
166	* PORT_AUTHORITIES	10		154		-39.79	0.000
167	* MARITIME_LAW	12		192		-50.94	0.000

* Negative keyness items such as *port_authorities* and *maritime_law* are not used in the linked keyword list because the linked keywords are concerned with positive keyness items.

Appendix C: Top 100 Keywords of a Law Corpus

N	Keywords	Freq.	%	RC Freq.	RC %	Keyness	P-Value
1	SHALL	16380	1.69	1730	0.06	35662.63	0.000
2	OR	16497	1.70	11040	0.37	15518.80	0.000
3	BE	17869	1.84	18179	0.61	10497.82	0.000
4	PARAGRAPH	2138	0.22	268	0.00	4464.96	0.000
5	CONVENTION	3449	0.36	1641	0.06	4200.24	0.000
6	PROVISIONS	2509	0.26	665	0.02	4152.81	0.000
7	ANY	4220	0.44	2905	0.10	3843.97	0.000
8	ACCORDANCE	1941	0.20	302	0.01	3840.35	0.000
9	PROVIDED	2321	0.24	861	0.03	3278.91	0.000
10	CERTIFICATE	1519	0.16	176	0.00	3227.68	0.000
11	SPACES	1285	0.13	84	0.00	3018.15	0.000
12	BY	9465	0.98	13629	0.46	2991.61	0.000
13	CARGO	2878	0.30	1721	0.06	2961.34	0.000
14	ORGANIZATION	2016	0.21	698	0.02	2952.35	0.000
15	REGULATION	1985	0.20	694	0.02	2892.48	0.000
16	REQUIREMENTS	2208	0.23	965	0.03	2837.68	0.000
17	CODE	1860	0.19	632	0.02	2749.68	0.000
18	PRESSURE	1586	0.16	376	0.01	2741.63	0.000
19	PACKAGINGS	924	0.10	0	0.00	2590.06	0.000
20	SHIP	4790	0.49	5187	0.17	2549.00	0.000
21	DATE	1506	0.16	407	0.01	2471.74	0.000
22	ADMINISTRATION	1505	0.16	408	0.01	2466.88	0.000
23	MEANS	1927	0.20	870	0.03	2426.36	0.000
24	MAY	4085	0.42	4146	0.14	2388.07	0.000
25	SUBSTANCES	1040	0.11	82	0.00	2374.74	0.000
26	TANK	1154	0.12	168	0.00	2323.06	0.000
27	REQUIRED	1865	0.19	862	0.03	2313.15	0.000
28	APPROPRIATE	1424	0.15	419	0.01	2252.94	0.000
29	TEST	1475	0.15	497	0.02	2189.31	0.000
30	SPECIFIED	1073	0.11	161	0.00	2143.00	0.000
31	NOT	6056	0.62	8339	0.28	2103.72	0.000
32	ANNEX	1043	0.11	176	0.00	2016.67	0.000
33	DANGEROUS_GOODS	922	0.10	97	0.00	1998.54	0.000
34	SUCH	3813	0.39	4183	0.14	1988.05	0.000
35	COMPETENT	897	0.09	116	0.00	1858.96	0.000
36	APPROVAL	1069	0.11	257	0.00	1837.82	0.000
37	PACKAGING	719	0.07	31	0.00	1774.59	0.000
38	PARTY	1386	0.14	627	0.02	1742.53	0.000
39	TANKS	916	0.09	177	0.00	1699.53	0.000
40	PROTOCOL	903	0.09	183	0.00	1649.56	0.000
41	MSC	1014	0.10	281	0.00	1646.43	0.000
42	CONTRACTING	853	0.09	177	0.00	1545.89	0.000
43	PORTABLE	638	0.07	34	0.00	1538.29	0.000
44	APPLICABLE	924	0.10	254	0.00	1505.39	0.000
45	AMENDMENT	705	0.07	83	0.00	1492.43	0.000
46	SUBSTANCE	628	0.06	44	0.00	1460.12	0.000
47	THAN	3136	0.32	3696	0.12	1456.14	0.000
48	COMPLY	567	0.06	22	0.00	1413.82	0.000
49	OF	45993	4.74	114530	3.86	1411.19	0.000
50	WHICH	5799	0.60	9376	0.32	1373.37	0.000

N	Keywords	Freq.	%	RC Freq.	RC %	Keyness	P-Value
51	AMENDED	684	0.07	102	0.00	1368.12	0.000
52	CONTAINING	643	0.07	79	0.00	1348.29	0.000
53	MATERIAL	1000	0.10	417	0.01	1321.54	0.000
54	RES	558	0.06	43	0.00	1278.61	0.000
55	PACKING	500	0.05	16	0.00	1267.77	0.000
56	PLASTICS	434	0.04	2	0.00	1191.98	0.000
57	APPROVED	829	0.09	298	0.01	1190.14	0.000
58	LESS	1411	0.15	1053	0.04	1186.91	0.000
59	REFER	502	0.05	32	0.00	1182.89	0.000
60	KG	514	0.05	39	0.00	1180.65	0.000
61	ARRANGEMENTS	776	0.08	251	0.00	1174.73	0.000
62	EQUIPMENT	1416	0.15	1087	0.04	1157.24	0.000
63	CONSTRUCTED	750	0.08	234	0.00	1154.95	0.000
64	ENSURE	1144	0.12	700	0.02	1154.08	0.000
65	PROCEDURES	891	0.09	387	0.01	1148.81	0.000
66	SUBJECT	756	0.08	250	0.00	1132.22	0.000
67	BALLAST_WATER	686	0.07	189	0.00	1116.47	0.000
68	SHIPS	3032	0.31	4122	0.14	1079.31	0.000
69	LEAST	1075	0.11	674	0.02	1062.42	0.000
70	OTHER	3434	0.35	5008	0.17	1050.16	0.000
71	IF	2337	0.24	2814	0.09	1045.55	0.000
72	MM	496	0.05	60	0.00	1043.64	0.000
73	UN	654	0.07	192	0.00	1035.43	0.000
74	WHEN	2335	0.24	2833	0.10	1030.94	0.000
75	AUTHORITY	1051	0.11	666	0.02	1029.28	0.000
76	PARAGRAPHS	439	0.05	31	0.00	1019.44	0.000
77	SOLID	544	0.06	104	0.00	1012.64	0.000
78	RECEPTACLE	361	0.04	0	0.00	1011.76	0.000
79	RECEPTACLES	365	0.04	1	0.00	1009.73	0.000
80	FIRE	799	0.08	360	0.01	1006.84	0.000
81	TYPE	1068	0.11	711	0.02	1001.48	0.000
82	APPLY	768	0.08	328	0.01	1000.40	0.000
83	LIQUID	507	0.05	81	0.00	995.37	0.000
84	EXCEPT	676	0.07	237	0.00	983.11	0.000
85	FLAMMABLE	388	0.04	13	0.00	980.06	0.000
86	EACH	1868	0.19	2047	0.07	974.44	0.000
87	DISCHARGE	603	0.06	171	0.00	969.30	0.000
88	INNER	406	0.04	24	0.00	966.31	0.000
89	TRANSPORTED	546	0.06	122	0.00	964.20	0.000
90	TEMPERATURE	572	0.06	147	0.00	957.94	0.000
91	SHOULD	2067	0.21	2505	0.08	914.27	0.000
92	MATERIALS	662	0.07	256	0.00	913.52	0.000
93	REFERRED	606	0.06	200	0.00	908.40	0.000
94	UNLESS	610	0.06	207	0.00	901.89	0.000
95	DEVICES	494	0.05	102	0.00	896.65	0.000
96	SPACE	759	0.08	378	0.01	895.04	0.000
97	CHAPTER	924	0.10	600	0.02	885.82	0.000
98	METAL	450	0.05	78	0.00	863.12	0.000
99	GUIDELINES	674	0.07	305	0.01	846.94	0.000
100	UNDER	2357	0.24	3189	0.11	846.92	0.000

Appendix D: Linked Keywords for Keyword Networks

ACCORDANCE (253)	rhumb_line, plies, hereunder, a-vi, measured, seagoing, accepted, corrosive, constituent, discharging, flashpoint, wood, action, duly, condition, sill, territory, contained, harmful, sealing, melting_point, apparatus, smoke, discretion, fusible, subjected, drums, below, instruments, ro, thereof, loads, without, protocol, entitled, those, insulation, inserted, listed, oxidizing, ros, manoeuvre, suitable, relevant, closed, seafarers, employment, after, react, account, openings, where, column, hazard, taken, shell, filling, the, fitted_with, carried, non-toxic, alloy, accommodation_spaces, fire, exemption_certificate, metal, conveyance, type, receptacle, ballast_water, flammable, meets, subject_to, packages, dispenser, damage, mass, under, inspection, equipment, loading, applies, safe, bulkhead, ullage, filled, oil_pollution, batteries, name, accessories, plugs, hour, list, alarm, hydraulic, escape, marking, pollution, moulded_depth, fire-extinguishing, appliances, vapour, notification, machinery_spaces, compartment, form, pesticides, deems, explosive, activated, commensurate, undertake, above, alternative, start-to-discharge, situated, dispensation, doors, precautions, righting, inadvertent, conform, sift-proof, trunks, potassium, address, recommended, longitudinal, closures, edp, appliance, load, integrity, walls, noise, nitric, piping_arrangements, engine-room, rupture, perform, security, steering_gear, cryogenic, particulars, stations, markings, calculated, stcw, prior_to, propulsion_machinery, covered, sealed, automatically, removable, self-heating, revised, surveys, complies, establish, decomposition, immersion, cleaning, disulphide, air_pipes, non-refrigerated, relative_density, installed, prewash, unprotected, exemption, navigational, bwmc, infectious, mawp, wearer, power-driven, deadweight, spaces, lights, asbestos, whenever, fuze, intervals, respect, quantity, normal, hereinafter, substances, languages, communicate, diameter, sides-cuttles, anti-fouling, occupational, circuits, exemptions, lifejacket, slop_tank, with, engaged, lockers, amending, globallast, purposes, arking, reference, valid, classification, disembarkation, solids, outer, gauging, passivated, adverse, freight_containers, approved, permeability, hatchway, satisfactory, withstand, protected, iapp, lifeboat, normally, preparation, multiplied, fitting, criticality, auxiliary_machinery, kpa, voyages, surveyed, manufacturer, danger, whose, subsection, pallet, biological, non-party, non-parties, if, issued, inspected, sediments, distress, plastics, standards, applicator, duties, verification, circulation, ship's, man-made
ADMINISTRATION (254)	closed, foam, suitable, g, before, accessories, devices, prevent, for, acep, defined, manifold, withdrawal, port_facilities, copy, where, permitted, adequate, person, minimum, on_board, non-hazardous, relevant, handling, fumigation, ullage, plies, liquids, vapour, examination, structural, cylinder, lubricating_oil, certificates, withstanding, inspection, whichever, described, passenger, incinerator, preparation, present, situated, sufficiently, above, protected, batteries, near-coastal, occupational, applicator, machinery, soon, the, stations, multiplied, radionuclides, constitute, haop, dangerously, subsection, segregation, residue, man-made, surfaces, hazard, ventilation, category, fibre, inspected, training, mass, lifejacket, fire_pumps, flexible, contracting, audible, damaged, intervals, sadt, request, wearer, stairways, non-conformities, automatically, labelling, unloading, compatible, flag, adequately, if, rescue_boat, bilges, adhesive, divisions, volumetric, msc, specimen, pesticides, exercising, kept, lining, source, subdivided, cartridges, approved, arrangements, amended, mf, appliance, detection, bwmc, given, nations, communicated, ingredient, furnish, supplement, bearing, pfsp, qualified, prior_to, complied, preventing, attachments, bags, unified, aqueous, hereto, within, denounced, revised, welded, waterproof, unattended, survival_craft, liferafts, attesting, waste, imdg_code, discharged, immunities, transported, notify, lifejackets, ui, except, all, fuze, overpacks, recommendations, inner, listed, entitled, power-driven, emergency_escape, every, inserted, open, cso, instructions, flange, parallel_of_latitude, ignition, condition, subjected, unintentional, normal_conditions, freeboard, pressure-vessel, ibc, competent, alarm, denunciation, particulars, quantities, resolution, either, performing, isolating_valves, proficiency, integrity, fitted, weather_deck, entries, accession, distress, megc, officers, required, recommended, stowage, coolant, specified, emergency, circ, reference, securing, peroxide, secretary-general, authority, sodium, liquid, ultraviolet, information, aerosol, valve, manually, name, acceptance, non-removable, self-heating, this, vents, ecx, adjacent_spaces, competence, special, valves, electrical, trimming, stairway, piping, tml, chapter, transverse_bulkhead, assigned, apparatus, receptacles, articles, machinery_space, security, regulation, procedures, maximum, parties, applicable, ships, authorized, paragraph, cargo, harmful, sealing, ship's, removable, sleeping, samples, withdrawn, deems, circumstances, permanent, filled, master, those, refilling

AMENDMENT (120)	liquid, thereafter, specified, identification, authority, survival_craft, sidelights, fulfil, metal, materials, comply_with, removable, intervals, seafarer, adequate, standards, use, self-reactive, party, precautions, manifold, machinery_spaces, duly, consignment, fire_extinguishing, due, package, man-made, necessary, determined, parties, present, sealing, deckhouses, mixtures, fire_alarm, for, dumping, radioactive_material, ingredient, may, purposes, machinery_space, constructed, minimum, shielding, category, than, ssp, provisions, prescribed, lifting, appliances, authorized, temperature, recommendation, above, address, except, anti-fouling, movement, health, pallet, such, fly, qualified, fittings, least, harmful_substances, lined, should, code, endorsement, expressly, organization, ducts, convention, carried, apply_to, un, conditions, persons, assembly, ships, reference, articles, fund, paragraph, lightest, accepted, pump, practicable, or, date, stairway, loading, facility, accordance, contracting, section, examination, specimen, external, not, megc, any, testing, on_board, class, by, refer_to, of, protocol, tested, segregation, renewal, be, shall, annex, after
APPROVAL (253)	gross_tonnage, cargo, gases, passenger, than, fissile, should, unless, guidelines, least, class, person, annotation, certificate, ratification, portable, kg, state, paragraphs, type, approved, closing_arrangements, peroxide, required, fitted_with, longitudinally, ventilation, exceed, constructed, metal, discharge, code, temperature, requirements, space, accession, piping, ultraviolet, means, transported, receptacle, filled, lifting, open, enclosure, radiation, carrying, non-removable, vertical, composite, centreline, alarms, sadt, specific, revised, complies, provision, routine, rupture, quantities, stairways, governments, g, standards, litres, pumps, megc, annexes, notwithstanding, manhole, detachable, performing, shielding, circuits, objectives, revalidation, lockers, iaea, whose, pesticides, diluent, conveyance, overpacks, freight_container, control_station, exercising, circumstances, vent, sub-systems, invasive, diameter, environment, examinations, acep, surfaces, frangible, operated, gross, pbz, harmful, packed, contained, seafarers, training, insulation, certifying, commensurate, fumigant, plate, precautions, ship's, fund, fibreboard, plies, litre, applies, does, copy, thereof, tables, electrical, fly, engaged, accidental, labelled, manifold, period, steering_gear, properties, radio_installation, parties, rhumb_line, stowed, flash, un, detection, control, minimum, damage, amended, survey, sprinklers, provisions, flammable, except, party, visible, toxic, concerned, watertight, recognized, radioactive_material, radionuclides, sound, appropriate, equipment, entry, ship, sodium, freeboard, oil_tankers, on_board, vents, prewash, this, chemical, globallast, met, adequate, placed, purposes, deems, emergency, ships, ducts, section, condition, packagings, appendix, used, certified, paragraph, organization, nozzles, sill, remanufactured, fire, solvent, reference, flexible, radio, liquefied_gases, consignment, loads, unpackaged, refer_to, valve, compatibility, seafarer, preparation, msc, account, containing, discharging, less, integrity, mass, either, intervals, pressure, specified, packaging, materials, distress, shut-off, protocol, criteria, contracting, test, applicable, capable, operation, navigating, following, such, competent, aluminium, other, chapter, mf, prescribed, accordance, loading, issued, res, noxious_liquid_substances, removable, cylinder, of, date, apply_to, material, annex, acceptance, subrogation, administration, filling, convention, covered, liquid, segregation, certificates, side-scuttles

CODE (253)	lifting, seagoing, fixed, hook, treated, discretion, assigned, fire-extinguishing, withdrawn, ros, exempted, degradable, attesting, loading, wastes, radioactive, incinerator, dermal, thereof, reception_facility, transmitted, filled, force, moira, sill, bulk_cargoes, unreasonable, amendments, hazards, having, trimming, watertight, toxicity, permitted, compartment, apparatus, identification, every, notification, those, into, aluminium, instruction, corresponding, non-parties, garbage, waterline, accepted, handling, permanent, alternative, listed, radio, below, due, hour, testing, vapours, weather_deck, revalidation, mg, impaired, undue, sound, surveyed, freeboard, authority, enter, cartridge, absorbent, drum, manoeuvring, latitude, recommended, heating, type, prohibited, closure, under, exhibit, glass, manufacturer, surveys, secretary-general, circulated, preferably, cylinder, furnishings, decomposition, comply_with, materials, aft, metres, fumigation, ambient_temperature, danger, sludge, approval, cargo_holds, annex, box, location, strength, refrigerated, conforming, non-flammable, surfaces, date, endanger, invasive, colour, amended, unit, pollutant, fire_damper, which, document, violation, constitute, sea-bed, pallet, portions, cooking, leakproofness, imdg_code, exceed, contents, stairways, defined, objectives, prevent, remain, self-heating, upwards, keels, conditioning, a-iii, wh, bq, rooms, sufficient, hazard, sediments, periodic, concentrations, column, waived, imsbc, sufficiently, purpose, languages, stowage, safety, respect, haop, walls, navigational, masters, pump-rooms, separated, communicated, closures, exemption, diluents, megc, transmitting, allowed, rocket, limitations, load, enclosure, switchboard, within, dsc, fresh_water, sift-proof, after, radioactive_material, subject_to, replaced, peroxide, substances, denunciation, organization, damage, spaces, following, transport, shell, pressure, explosive, un, letters, tanks, bulk, applicable, adequate, discharge, on_board, specified, accession, criteria, chapter, tests, protected, other, min, section, individually, halon, maintained, adjacent, declaration, dust, ignition, confinement, biological, room, mf, take, satisfaction, draught, source, health, behalf, gauging, responsibilities, stairway, security, request, nominated, pigments, inf, facilitate, laws, globallast, environment, qualified, purposes, medical_first_aid_guide, edp, self-reactive, sealing, ro, forth, certificate_of_fitness, competence, residue, sides-cuttles, fund, inter, arranged, examination
COMPETENT (255)	litre, containers, competence, standard, applies, persons, authorized, receptacle, metal, gases, adopted, special, subject_to, paragraphs, applicable, substance, regulation, portable, kg, ensure, than, which, openings, hour, doors, having, compartment, signature, toxic, protected, fixed, state, certification, stowed, on_board, haop, column, shell, capable, denunciation, arrangement, manner, flange, satisfaction, suitably, litres, radiation, permissible, navigating, see, cartridge, revised, penetrations, animals, mg, upwards, particulars, bulkhead_deck, complied, lithium, embarkation, stairways, pyrotechnic, edp, whenever, an, bwmc, load, appliance, entries, alternatively, extinguishing, water-spraying, min, diluent, transportable, modifications, damaged, dermal, withdrawn, deckhouses, discharges, enclosure, lifting, adverse, circumstances, instruments, emitters, violation, drum, centreline, hereinafter, preparation, distress, radionuclide, contaminated, visual, latitude, impaired, mcconnell, machinery_space, listed, harmful, entitled, contained, this, determined, enters, ignition, oxygen, measures, proper, either, recognized, packing_instruction, escape, self-reactive, inserted, seafarer, aluminium, audit, certified, security, intervals, languages, classification, accedes, parties, intermediate, radioactive_material, nobob, means, quantities, vent, ventilation, steel, deformation, packagings, machinery, ship, administration, following, survey, marking, fastened, biological, alarm, steering_gear, self-heating, cargo_units, plastics, res, consignment, exercising, attesting, annex, carriage, specified, gross, decks, request, fissile, accepted, amendments, ro, ratification, appliances, unpackaged, segregation, non-refrigerated, entry, exceeding, detectors, sufficient, carrying, watch, pursuant, prevent, circ, cargo, evacuation, notify, requirements, account, man-made, equally, ui, permitted, visible, radiotelephone, contracting, deems, liquefied_gases, schedules, closures, ships, objectives, for, relief, name, un, decides, inspection, passenger, under, conform, recalling, preclude, removable, referred, properties, danger, temperature, temperatures, dangerous_goods, isu, duly, pressure, radionuclides, battery, c-weighted, oil_tankers, fibreboard, pesticides, certificate, accordance, provisions, watertight, provided, code, remanufactured, tank, space, vents, person, except, acceptance, liquid, filled, chemical, issued, constructed, used, less, approval, such, package, after, constituted, damage, msc, least, containing, packaging, each

CONVENTION (255)	mcconnell, man-made, address, activation, assembly, with, objection, proof_strength, control, approving, vents, due, thermal, fusible, side-wall, substances, stability, carriage, placement, fire_extinguishers, atmospheric_pressure, for, account, near-coastal, manoeuvre, country_of_origin, maximum, arranged, equivalent, danger, recognized, dangerous_goods, established, inspected, inadvertent, sill, attained, ambient, biological, outside, residue, meets, representative, physical, hydride, anti-fouling, emergency, definitions, appliances, protected, oil_tankers, present, potassium, mf, unaccompanied, formulations, regard, wastes, discharges, plastics, radio_installations, determining, relief, air_pipes, casings, fire_detection, assigned, oil_pollution, deckhouses, sadt, two-thirds, standard, certification, pump, liquid, reasonably, stacking, sulphide, application, detection, nozzles, issued, associated, megc, manufacturer, annotation, withstand, nominal, less, load, continuous, fuse, exceed, infectious, filled, undertake, description, radiation, manufacture, comply_with, non-fixed, officers, manufacturer's, occupational, placard, inboard, chloride, substance, information, decides, behalf, propelling, complies, personnel, incineration, cooking, chapter, fibreboard, baggage, auxiliary_machinery, ui, fumigant, schedules, stcw, ensuring, free_surface, denounced, relevant, corresponding, trimming, carry, consignment, state, gross, ships, piping, leaking, accidental, transmitted, embarkation, cell, fund, pollutant, required, shut-off, gases, instruments, vent, pressure-vessel, wagon, samples, receipt, plies, outlets, engine-room, health, taking, sodium, re-use, supplementary, excess, lifejackets, external, intervals, vehicle, liquefied_gases, radioactive, undue, polymerization, edp, transmitting, section, if, satisfactory, refrigerated, whenever, fire_alarm, ship?, outfits, open, when, burst, ies, appliance, unit, circ, machinery_space, impracticable, hexafluoride, emergency_escape, into, multiplied, survey, fitted, smoke, mandatory, subjected, msc, manhole, this, placards, material, specimen, purposes, inspection, multiwall, appendix, compartment, deformation, witness, rhumb_line, expiry, fuel, preparation, cargo, machinery, permitted, deck, non-party, self-heating, quantity, stairways, specific, materials, lining, acceptance, endorsement, uranium, flange, determined, mass, transportable, satisfaction, arrangements, placed, constructed, linings, a-ii, kg, voyage, measured, particulars, res, adjacent, quantities, plate, assumed, withdrawal, lifejacket, openings, water-spraying, sides-cuttles, used, aquatic, loads
MAY (255)	particulars, governments, consultation, pathogens, fire_detection, cushioning, ro's, flooding, exposure, committee, virus, a-iii, weathertight, integrity, normally, closures, establish, designed, provision, side, markings, switchboard, accommodation, heat, audible, fire_extinguishing, conformity, conf, non-hazardous, loaded, columns, sediments, port_facilities, normal, quantity, plugs, restricted, mechanically, expiry_date, engine-room, stairways, solas, sbt, decks, due, annotation, unloading, sub-paragraph, damaged, non-conformities, second_engineer, side-wall, vertical, removal, acute, fuze, ies, issuing, correspond, insulated, draught, recommendations, specific, fastenings, jerricans, crustacea, attached, non-refrigerated, manually, hypochlorite, limits, effected, inert_gas_systems, prohibited, plies, pump-room, casing, affixed, g, functions, fishmeal, litre, guideline, assistance, diluent, crude_oil_washing, country_of_origin, non-parties, expelling, undersigned, longitude, ghs, automatic, iopp_certificate, bottom, melting_point, renumbered, storage, equivalent, fitness, into, fire-extinguishing, fixed, expiry, duties, certificates, saunas, tubes, ion, control_station, for, a-v, packing_instruction, liquefied_gas, medical, electrical, whichever, effluent, inadvertent, fireworks, category, spontaneous, criteria, emit, space, each, inspection, store-rooms, metal, amendment, protocol, walls, receptacle, solid, packing, symbols, capable, stowed, machinery, hazard, replaced, ratification, emergency, particular, longitudinally, foam, packages, venting, survival_craft, area, sco, designated, a-vi, hazardous, non-metallic, apply, glued, reasonable, hazards, biphenyls, oxygen, tank_top, decides, cyanide, mhb, dusts, waterline, discharged, officer, reference, ship's, apparatus, cube, jacket, contains, ssp, ballast, voyage, definitions, before, enclosed, attesting, non-cohesive, waived, plastic, bridge, condition, subjected, including, toxicity, eparated, expressly, power_supply, cargo_holds, ammonium, textile, compounds, depleting, communicate, carry, exemptions, an, epirb, pallet, cooling, http, bulk_cargo, heads, inert, documentation, vent, stopping, temperatures, general, familiarization, gastight, document, available, inorganic, wreck, stations, remain, resistant, combustion, disembarkation, adhesive, aerosols, nuclide, cm, trays, ambient, colour, lubricating_oil, receipt, hexafluoride, fitting, linings, uptake, certifying, excess, foam_solution, sanitary, inhalation, mark, approving, source, undertake, recycled, diseases, height, spare, casings, depositing, on_deck, co-operate

ORGANIZATION (255)	fly, device, capacity, entitled, deemed, fibreboard, liferaft, references, passenger, wagon, inserted, articles, regulations, condition, proper, obstructions, recognized, sbstta, filled, insulation, steel, for, name, radionuclide, attached, additives, flammable, type, accumulator, implementation, design, width, coolant, combustion, mm, thereafter, comply_with, two-thirds, under, circumstances, considered, temperature, refrigerated, radionuclides, cylinder, lubricating_oil, recommendation, fire_extinguishing, ullage, periodical, water-spraying, reception, res, aerosol, objection, miscellaneous, emergency_escape, material, nobob, tml, lining, behalf, transportable, dose, authorize, labelling, cso, attesting, vehicles, composition, adjacent_spaces, gaskets, embarkation, taking, radiocommunications, endorse, intervals, deckhouses, quantities, closed, pfsp, prior_to, the, shell, radio, without, compartment, vapour, those, carriage, maintained, assigned, hoses, volumetric, reference, longitudinal, extinguishers, ventilation, special, chief_engineer, minimum, melting_point, acceptance, mass, react, stations, near-coastal, removable, lights, complies, pursuant, evacuation, entries, sufficient, freight_containers, manually, port_state, label, code, certificate, certification, openings, resolution, issued, state, radioactive, hazard, arrangements, less, portable, ships, parties, space, fire, medical_first_aid_guide, candidate, lifting, determine, particulars, suitably, standards, edp, environment, objectives, weather_deck, absorb, pesticide, casing, quantity, deems, stcw, request, labelled, consignment, vehicle, recommended, external, carry, latitude, as, non-flammable, loads, inside, use, leakproof, voyage, flexible, mark, twelve, freeboard, furnishings, sodium, aqueous, discharging, address, survival_craft, combination, pressure-vessel, multiwall, flashpoint, sprinklers, pollutant, subjected, every, all, samples, seafarer, verifying, competence, anti-fouling, used, mixture, performing, containers, mechanical, signature, on_board, wastes, gases, circuits, specific, present, chemical_tanker, regulation, suction, into, spaces, marking, auxiliary_machinery, sound, reconstituted, oil_tankers, duties, not, machinery_space, withstand, sadt, declaration, determined, hydrogen, approval, matters, loading, megc, pump, ensure, satisfactory, bar, defined, sealed, globallast, rupture, gross, endanger, fulfil, procedures, certifying, machinery, annual_survey, health, compatibility, test, surveyed, chapter, vents, freeboards, provided, ecx, date, amendment, nozzles, unpackaged, duly, equally, drum
PACKAGINGS (253)	valves, carriage, having, constituents, contents, certification, capable, periodic, radioactive_material, gases, denunciation, angle_of_repose, exceeding, garbage, doors, compliance, filled, width, suitable, state, freeboards, complying, due, sbt, atriums, restricted_areas, conform, checked, freight_container, a-i, owner, extent, bilge, subsection, passivated, bar, malfunction, implementation, sub-section, properties, wh, mf, quantities, schedules, pathogens, pesticides, res, manufactured, loads, covered, breathing, recommended, required, notifications, centreline, washing, composite, heated, seafarersrights, specified, harm, symbol, surveys, rescue_boats, pursuant, arrangement, subparagraphs, freight_containers, equal, residue, specifications, allowed, endeavour, referred, labelling, insulation, toxicity, fulfilled, except, every, including, subjected, control, charge, medical, following, special, salvage, fly, compartments, excepted, audit, alternative, roof, other, purposes, communicate, if, dioxide, performing, non-opening, salvor, aqueous, certified, definition, this, hoses, trimming, authority, ballast_tanks, flammable, equipment, substance, applicable, familiarization, rhumb_line, sea-bed, paragraphs, approval, ratification, unless, competent, such, ballast_water, acceptance, tanks, refer_to, necessary, exemptions, meets, door, cellulose, globallast, balconies, enters, organic, recognized, electrical, survey, vapours, horizontally, apply, engaged, closure, metres, revised, placed, non-refrigerated, water-resistant, fibre, indicated, air_pipes, nominated, heating, separated, navigating, edp, purpose, crude_oil_washing, valve, shell, cooking, welded, non-conformity, aerosol, cargoes, prevent, preclude, store-rooms, tests, force, packages, passenger, segregation, solid, insulated, flange, shielding, means, removable, dangerous_goods, amending, moisture_content, machinery_spaces, section, load, davit-launched, list, ibcs, temperature, organisms, provided, leakage, kpa, ventilation, bulkheads, non-metallic, damage, instructions, steel, with, particulars, of, code, any, amended, subject_to, vents, cell, freeboard, foam, circ, constructed, lithium, sub-systems, gross, accordance, haop, manual, arrangements, requirements, machinery_space, account, pressure, deposited, inter, un, recycling, additional, liquid, trays, substances, parties, ammonium, tested, drainage, surveyed, harmful, material, by, approved, liquids, harmful_substances, lightest, than, mm, test, may, issued, least, facility, ballast, receptacle

PROVIDED (254)	issued, a-weighted, survival_craft, pyrotechnic, temperature, recesses, flashpoint, alarm, recognizing, conform, rooms, side-wall, plywood, fire_dampers, training, materials, list, bs, a-iii, manual, batteries, subdivision, venting, connected, adjacent, without, packing, mixtures, medical, side, accommodation, piping, bulkhead_deck, bulk_cargoes, gross_tonnage, machinery, signature, water, switchboard, when, recognized, contracting, aquatic, valve, constructed, quantities, moisture_content, radioactive, designated, functioning, concerned, solids, audit, compartments, units, this, declaration, intervals, ro, mercuric, testing, adequate, doors, drums, equivalent, emptying, flag, globallast, exceed, openings, contained, article, relevant, acid, infant, due, survey, respect, instructions, foam, engaged, articles, residues, mass, consignment, air-cushion, embarkation, acceptance, steering_gear, fixed, watchkeeping, examination, replaced, non-commercial, halon, taking, metal, chapter, seafarers, subparagraph, rigid, class, maximum, resolution, defined, explosive, fittings, assembly, oxygen, subject_to, stowed, assistance, voyages, present, ballast, minimum, obligations, specification, oil_tanker, gross, particles, sewage, foregoing, circulation, biological, invites, refrigerant, fitting, electrically, fibre, unified, surfaces, passivation, temperatures, tween-deck, synopsis, acep, evacuation, cleaning, ullage, sprinkler, damaged, caps, nos, acceded, cyanide, centreline, mepc, digits, sidescuttles, double_bottom, displayed, alternatively, vertical, accidental, document, container, dispensed, cube, emitters, sound, animals, measured, aqueous, fans, member, unintentional, distress, plugs, pallet, rudder_stock, biphenyls, additives, stop-valve, propellant, vents, visible, enclosures, notifications, avoid, oil-fired, determining, linings, port_facilities, on_deck, preparation, undertake, non-flammable, inoperative, lifting, amend, plastic, height, mercury, description, meets, baggage, chemical_tanker, refrigerated, meteorological, correspond, deformation, box, reconstituted, lubricating, decomposition, deems, radiocommunications, respiratory, effective, leakproof, cmnd, trimming, stacked, load_line, fire-fighting, litres, hereunder, address, freight_containers, immersed, gauging, issuing, drying, complies, coolant, surveys, rescue_boats, megc, balconies, inhibitors, trays, satisfactory, indicating, g, exercising, limitations, manned, conditioning, indicate, textile, automatic, constituents, unattended, witness, ventilated, transmit, non-removable, cso, sounding_pipes, cargo_hold, load_lines, bar, ecx
PROVISIONS (254)	ballast_tanks, iso, measured, operation, cooling, manufacturer, hazards, charge, paragraphs, moira, ballast_water, prior_to, machinery_space, specification, non-hazardous, wearing, glass, xa, pressure-vessel, regard, following, facilitate, psy, bulkhead, obstructions, packages, cylinders, performing, pyrotechnic, tml, switchboard, packed, non-flammable, salvage, reasonably, liferaft, duly, escape, hazardous, deballasting, within, quantities, cube, package, pathogens, condition, smoke, flashpoint, drums, ballast, inter, diluent, objectives, corresponding, ds, manhole, zinc, quantity, iii, renumbered, deemed, copy, ignition, thickness, watertight_doors, with, procedures, oil_tankers, balconies, width, maximum, enters, inserted, maintained, manufacturer's, expires, mentioned, authority, bs, radio, ships, before, thereto, extend, identification, rigid, properties, ibcs, nitric, embarkation, unloading, permit, prescribed, commensurate, attached, intervals, marked, a-vi, resolution, accidental, discretion, carbon, notifications, constitute, whistle, organisms, survival_craft, detonation, corrosion, loaded, height, exceed, space, self-heating, refrigerant, ibc, immersion, pfso, manually, open, suitable, amend, load_line, oil_pollution, on_deck, system, protection, withdrawn, application, criticality, carrying, automatically, arrangement, stored, properly, gauging, sub-systems, manner, re-use, pbz, shipboard, g, equal, metal, specifications, residue, alarms, bulk, mepc, radioactive, frangible, chlorate, chapter, navigating, edp, seafarersorganizations, depositing, fireworks, nominal, sodium, discharges, refrigerating, machinery, assigned, indicate, nearest, referred, specific, centreline, ui, bilges, conformity, drills, hydroxide, fastenings, persons, defendant, gases, exemptions, conveyance, a-weighted, deems, pesticide, protected, fire-extinguishing, applicator, ls, complying, infectious, continuous, tie-down, described, stowed, familiar, fuze, ingredient, min, whichever, implementation, eparated, accession, ullage, lifejacket, where, pursuant, subsection, double-side, exceeding, washing, crude_oil_washing, closed, equivalent, dumping, disembarkation, receipt, irrespective, deck, closure, radiation, consignor, on_board, testing, loads, periodic, bursting, certification, gas-freeing, free-fall, effective, verifications, metres, cells, hazard, cylinder, adjacent_spaces, the, insulation, piping, recycling, solid, cargoes, fund, fibreboard, electrical, temperature, constituted, fitted, person, packing_instruction, manual, containers, safe, marking, bulkheads

REGULATION (254)	concentration, water-spraying, undue, territory, suction, without, shell_plating, cso, watch, enters, uranium, with, properly, acid, discharges, inform, evidence, damage, reference, freight_container, flashpoint, plugs, gauge, vicinity, preclude, anti-fouling, discharging, qualifications, removed, inhalation, under, discretion, appropriately, persons, before, aquatic, authorized, nitrates, oxidizing, thereof, recommendation, ullage, item, accepted, bulkhead, dispersible, surfaces, training, flash, tissue, containers, mixtures, charge, inter, inspection, whichever, capacity, acceptance, bulk_cargoes, ro, operable, form, isu, iaea, high-expansion, include, trimming, additives, circumstances, listed, satisfy, those, samples, inboard, moira, endorsements, deflagrating, material, oil_pollution, frangible, secretary-general, organic, intact, applicator, condition, melting_point, officers, fulfilled, insulation, steel, rude, chloride, mark, replaced, continuous, dust, load, auxiliary_machinery, closures, cyanide, infectious, stowed, appliance, guideline, constructed, protocol, righting, movable, protectors, cartridge, linings, composite, supplement, sulphide, transmit, width, normally, meets, end-wall, a-i, packages, temperatures, slop_tanks, purpose, freeboard, part, connexion, sufficient, indicating, non-flammable, power-operated, db, oprc, establish, consignor, intended, granular, fitted_with, lighting, amended, cell, detection, labelled, reg, haop, whenever, recommendations, cartridges, indicate, competent, relevant, extinguishing, security, lithium, evacuation, systems, recesses, manoeuvring, subsection, leakproof, dumping, accession, emergency_escape, perform, waterproof, shipownersand, vii, prejudice, sprinklers, sealing, outside, trained, noec, garbage, languages, system, general_purpose_container, responsibilities, ballast_water, alternatively, take, propulsion_machinery, sufficiently, oil_record_book, absorbent, centreline, walls, cdb, within, modified, column, text, entry, multiplied, particulars, special, one-third, oil_filtering, drills, request, fire_control, stairways, loading, rooms, adhesive, swap_bodies, gauging, carbon, gases, practicable, state, permeability, approving, above, fittings, testing, into, deckhouses, expelling, seafarers, employment, boiler, doors, cargo_holds, amend, periodic, synopsis, plan, defined, contracting, maximum, refrigerated, manifold, mandatory, ventilation, ibc, watchkeeping, flooding, rigid, weather_deck, non-party, ships, marks, cushioning, laws, hydrogen, placarding, maintained, flange, globallast, dry, performing, sound, withstanding
REQUIRED (255)	uranium, fulfilled, multiplied, trimming, peroxide, batteries, salvage, readily, suction, date, inert_gas_systems, iv, engaged, approval, specimen, deposit, packagings, sample, cylinders, occupants, marking, receptacles, lever, constituted, units, constructed, period, sediments, competence, receptacle, acid, disembarkation, temperature, ds, chapter, effective, substances, inner, arrangements, audible, intervals, either, classified, kpa, rhumb_line, exceeding, accessories, instruction, biological, periodic, smoke, the, hazardous, organisms, aerosols, power_supply, proper, confinement, certificates, assigned, carriage, measures, placarded, non-metallic, compartments, examination, aluminium, transport, steel, machinery, copy, accepted, shell, sprinklers, otherwise, plies, master, assistance, co-operate, passenger, overpressure, ignition, remanufactured, ventilation, following, iso, precautions, flashpoint, approves, exceed, entry, tests, listed, closed, defined, radio, packages, liquids, before, fitted, vapours, necessary, treated, adopted, flexible, voting, source, two-thirds, pursuant, excess, luggage, cell, decomposition, water-resistant, whose, bilges, boilers, oil_tanker, notifications, considered, loads, detector, lights, location, walls, stacking, rocket, proficiency, reg, automatic, angle_of_repose, free-fall, preventing, sealing, stores, chronic, fibre, structural, auxiliary_machinery, revised, recharging, available, limits, unattended, db, freeboard, documented, composite, movable, marpol, functions, non-party, edp, labelling, sulphide, surface, drainage, as, cyanide, securing, transverse_bulkhead, fire_extinguishers, authorize, passivation, withstanding, covered, stations, nominal, fans, csc, double_bottom, supra, follows, moisture_content, within, seafarers, employment, particulars, marks, danger, deballasting, non-cohesive, aqueous, unit, refrigerated, bulkhead_deck, unsafe, properly, bags, g, steering_gear, properties, distress, bulk_cargo, vents, attached, react, rigid, sbstta, a-iii, recalling, bursting, mcconnell, meteorological, see, reporting, stop-valve, preparation, dust, continuous, sound, trays, volumetric, leak, exemption, overpacks, bwmc, emergency_escape, undertake, unintended, sift-proof, situated, discharge, each, cargo_units, present, corrosion, pyrophoric, regard, space, fittings, recommendation, on_board, calculation, render, transported, accession, circuits, less, decides, non-conformities, width, wastes, fire, paragraphs, epirb, amended, plastics, name, acep, issued, vent, cylinder, damage, class

REQUIREMENTS (255)	arranged, lithium, meteorological, overpressure, security, above, melting_point, apply_to, pfsp, bulk, undertake, cleaned, instruction, deckhouses, cube, seafarer, accumulator, non-parties, authorize, rigid, control, indicated, oil_pollution, decks, permit, solas, tests, machinery_spaces, mechanical, contain, end-wall, intact, inorganic, adequate, ro, refrigerated, sprinklers, dangerous, unloading, soon, inert_gas_systems, qualified, ro-ro, deposit, applicator, ambient_temperatures, solids, unduly, fire_damper, instructions, baggage, connected, assistance, vapours, neutron, durably, wastes, attachments, filled, thereto, unit, escape, assembly, without, evidence, mcconnell, signatures, lining, gauge, moira, corrosive, liferafts, connexion, packing_instruction, valid, litre, breathing_apparatus, co-operate, cylinder, flashpoint, standard, flooded, batteries, distance, reasonably, fire-extinguishing, pesticides, engaged, territory, thermally, readily, loaded, supplemented, capacity, peroxide, cargoes, bulk_cargo, physical, dry, fumigation, disposal, electrical, flag, this, fusible, packaged, aquatic, globallast, distress, information, before, width, chemical, metres, identifiable, leakproofness, copies, devices, guidelines, inspection, boxes, under, emergency_escape, labelled, calculation, decides, body, inner, sodium, fuse, removable, effected, freeboards, arrangement, muster, draught, expelling, outer, cell, emergency_response, a-ii, parties, met, enclosure, removal, subject_to, maximum, placed, dermal, accedes, diluents, approval, adhesive, possessing, trunks, capacitors, centreline, recharging, surveyor, crude_oil_washing, container, area, washing, power-driven, tank, reasonable, malfunction, annex, stations, denounced, furnishings, inspected, navigating, aqueous, type, inboard, heated, non-opening, vehicle, portable, powder, internal, wiring, routine, trifluoride, outflow, substance, haop, fire-fighting, placard, refrigerant, ibc, inhalation, pyrotechnic, switchboard, deflagrating, suitable, circ, column, volumetric, purpose, surfaces, correspond, consist_of, acute, withdrawn, fibre, criticality, normally, bwmc, certification, conform, reception_facility, equal, undertakes, bottom, on_board, heat, vents, db, preventing, adjacent, liability, officers, breathing, laws, satisfactory, criteria, quantities, exceed, contents, transmitting, sub-systems, cooking, bilge, parallel_of_latitude, machinery, drying, ballast, deemed, accidental, substances, iaea, deems, transport, taken, carriage, passengers, competent, state, lifting, vapour, audit, linings, operable
SHALL (254)	internal, no, treated, burning, establish, operable, hatches, effective, guidance, a-v, toxicity, nobob, segregated, horizontal, propelling, slop_tanks, apparatus, nozzles, cofferdams, pollution, undue, marked, ships?, digits, affixed, moulded, perchlorate, capacitors, general, applied, sadt, salvor, interpretations, annotation, described, steel, criteria, superstructures, breathing_apparatus, connexion, preferably, vii, specimens, sanitary, commensurate, refer, auxiliary_steering_gear, dose, wearing, general_purpose_container, sheets, decontaminated, drop, circulated, vehicles, metals, manufacturer, annual_survey, sprinklers, bearing, respects, iv, dirty_ballast, gastight, deadweight, checking, fissile-excepted, adversely, criticality, unified, boiler, eparated, permeability, solvent, allowed, ls, decease, longitude, inflated, verify, hose, given, thermally, circulate, regard, exempted, masters, consisting, fore-side, flammability, calcium, lsa, adsorbed, tank_top, purging, circuits, seafarerstraining, davit-launched, approves, cryogenic, requirement, carriage, correspond, undertakes, familiarization, second_engineer, device, packing_instructions, applies, design, copies, leakproofness, inside, relieving, effect, surface, pipes, nitrates, permanent, shell_plating, inert, vent, document, associated, deballasting, depleting, deformation, container, double_bottom, columns, invites, malfunction, entrances, hexafluoride, oily, movement, outside, aggregate, sts, prevention_of_pollution, frangible, imdg, oily_mixtures, latitude, programmes, damage_stability, imsbc, inhibitor, forming, sounding_pipes, deflagrating, slop_tank, uptake, liquefied_gas_tanker, addition, capacitor, indicate, pbz, passivation, regarding, control_stations, skin, oxide, delegated, portions, application, radiocommunication, specific, located, manner, divisions, hydrostatic, steps, deflagrate, immunities, ventilating, accede, unintentional, cause, nuclide, threaded, liquefied_gas_tankers, functions, furnish, cartridges, reasonable, log-book, trained, sprinkler, react, perpendicular, inert_gas_system, stability, installed, see, vehicle, enter, heating, word, corrective, schedules, absorbent, oil-fired, installations, closure, biopollution, inert_gas, nitrate, determined, radar, survival, accompanied, furnished, certify, cso, defect, thorium, limit, skids, ambient_temperatures, peroxides, invaders, phlegmatizer, hold, unreasonable, keel, hook, symbols, ds, fuze, ambient, lsa-i, representative, oversight, statutory, non-conformity, length, all, contaminated, rso, waterline, liferafts, power_supply, supplemented, accommodation_spaces, imdg_code, waste, ammonia, oxygen, non-metallic, valid, sub-paragraphs, plastic, xa

SPACES (253)	relief, accessible, valve, activated, adverse, departure, dumping, candidate, movable, refrigeration, meets, languages, containers, liquid, unless, fittings, deposit, gmdss, partitions, validity, establish, inner, deems, surveyed, visual, referred, subparagraph, depositing, side, organisms, deficiencies, plies, bags, damage_stability, cargo_units, g, ships, records, unsafe, xa, handling, assumed, cooling, gross, heat, bilge, cylinder, proficiency, depositary, otherwise, fly, impair, pallet, iso, litre, centreline, sub-systems, periodic, effects, haop, sufficient, class, pipes, seafarersemployment, effect, granules, denunciation, capable, inflated, alterations, inert, accession, resolution, ratification, liquids, person, shell_plating, respects, foam, lubricating_oil, present, maximum, after, appliance, individually, checking, withstanding, under, constituent, loads, mass, samples, width, doors, amend, tests, trained, respiratory, radioactive, state, remanufactured, distance, shell, pfso, ingredient, vents, load, compartment, valves, acceptance, decease, nitrates, prior_to, measured, aqueous, dangerous_goods, righting, survival_craft, surveys, door, flashpoint, consignor, internal, acute, annex, adjacent_spaces, fire-fighting, notify, carbon, available, removable, material, exempt, radio, subdivided, subparagraphs, surfaces, arrangement, listed, specified, power-operated, consisting, responsibilities, radiocommunication, hypochlorite, volumetric, accordance, decomposition, specimen, normal, flame, cryogenic, this, divisions, reasonable, substances, duct, manhole, boxes, units, consignments, conforming, combination, power-driven, dry, synopsis, smoke, packed, sewage, humidity, min, dust, deposited, alternatively, intermediate_survey, imsbc, compatible, slop, indicate, revised, res, nitrogen, drums, jerricans, suitably, adversely, plastics, equipment, code, watertight_doors, vent, assigned, decimal, months, provision, decks, list, self-heating, impaired, assistance, approved, overpressure, gases, refrigerated, second_engineer, rhumb_line, inorganic, satisfaction, megc, recommended, automatically, residue, closed, ballasting, control, intended, cease, noise, suitable, defined, instruction, circ, specific, transported, non-cohesive, portable, concerned, gauge, standard, fire-extinguishing, reservation, above, safe, opening, harmful, inter, distress, particulars, applicator, continuous, normal_conditions, no, limits, water, appropriate, test, electrical, carriage, seafarers, associated, containment, except, vertical_zone
SPECIFIED (253)	transfer, kg, liability, less, rudder_stock, nitrocellulose, rescue_boats, inflatable, centreline, criticality, requirement, appliances, machinery_spaces, flag, solids, linings, date, withdrawn, filled, following, flooding, damage, name, conditions, determined, segregated_ballast, denounced, cleaned, standard, peroxide, heating, subparagraphs, manifold, hour, radioactive_material, inspected, freeboards, shipboard, unreasonable, substances, stowed, capable, near-coastal, on_board, dioxide, issued, accession, declaration, width, emitters, arrangements, circ, fibre, exceed, foam, unprotected, residue, fitted_with, closed, spaces, remanufactured, side, having, valves, surveyor, cushioning, bulk, on_deck, adequate, dispensation, instrument, compartment, periodically, shell, potassium, pathogens, enters, quantities, surveyed, constructed, loading, packagings, multiwall, valve, referred, organic, liquid, radio, annexes, arrangement, displayed, gas-tight, limits, deficiencies, seafarer, classification, deemed, approved, emergency_escape, prewash, min, ro, receptacle, portable, inert_gas, suction, discharge, part, whenever, inner, amended, decides, tween-deck, iaea, personnel, receptacles, gauging, material, non-conformities, temperature, insulation, amendment, responsibilities, protection, lifting, contained, mass, under, distress_alerts, glass, months, electrical, furnishings, discretion, stairway, flange, inspection, appliance, organizations, exemptions, category, surveys, assumed, sulphide, located, concentrates, abandon_ship, violation, deems, internal_combustion_engines, covered, designated, list, adverse, medical, flame, recognized, each, res, substance, relevant, paragraphs, competent, ventilation, cargo, annex, rhumb_line, sides-cuttles, carbon, refer_to, should, code, un, adjacent_spaces, thermally, ships, person, necessary, fitted, administration, metal, suitable, proper, garbage, freeboard, freight_container, entitled, measures, water, toxic, subparagraph, containers, wastes, compliance, subdivided, extinguishing, purposes, within, given, particulars, reference, listed, machinery_space, globallast, free-fall, officer, segregation, sufficient, passenger, haop, depositary, denunciation, amend, noxious_liquid_substances, periodic, vents, washing, refrigerated, systems, ducts, ensure, provided, procedures, gases, automatically, plastic, satisfaction, any, self-heating, candidate, health, tanks, pipes, may, required, appropriate, hydrogen, edp, equipment, chemical, than, which, flexible, materials, ratification, space, fibreboard, duties, certificates, removable, every, cargo_holds, freight_containers

SUBSTANCES (254)	watertight, paragraph, iso, definitions, leakage, examination, accordance, convention, functioning, alarm, unsymmetrical, should, periodic, referred, sufficient, stowage, mercuric, taken, back-off, receptacle, ibc, capable, filling, type, ventilation, ratification, gases, ibcs, special, minimum, package, metal, foam, prevent, end-wall, on_board, fireworks, eparated, specified, appropriate, water, dangerous_goods, containing, above, msc, packing_instruction, ship, spaces, applies, filled, required, corr, ships, instrument, constructed, bounding, paragraphs, if, tank_top, watch, passenger, packing, less, explosive, contain, equipment, noxious_liquid_substances, seams, sides-cuttles, vents, centreline, flange, revised, corrective, transmit, symbol, cargo_holds, mechanical, preparation, ambient_temperature, preclude, inspected, non-flammable, box, level, furnish, exemption, edp, suitably, appliance, overpack, a-i, continuous, load, water-resistant, unpackaged, acceptable, load_line, detectors, spare, consignor, placed, application, garbage, periodical, assumed, rupture, walls, radiocommunication, overturning, fusible, cube, electrically, tables, boiler, purging, treated, unreasonable, alternative, trimming, degradable, accumulators, iopp_certificate, diluent, electro-technical, transfer, statutory, pyrophoric, visual, limit, overpacks, deleted, fitting, stabilized, compatible, authorize, communicate, performing, source, protectors, relating, radio, listed, rigid, survival_craft, thereto, permit, tabulated, particulars, obstructions, rooms, subjected, non-conductive, intervals, certified, respect, responsibilities, safety, discharged, a-vi, oil_tankers, audible, determined, steel, audit, ballast, survey, self-reactive, oily, organic, packed, wood, harmful, code, openings, requirements, following, pressure, maintained, manual, manholes, reservation, safe, circumstances, purposes, outer, hydrogen, specimen, residue, remanufactured, biphenyls, notify, deadweight, hazards, emergency_escape, packaging, extinguishing, plastics, material, kg, enclosed, forth, devices, containment, transport_document, list, means, flammable, marked, co-operation, materials, escape, drum, pump-rooms, indicated, trays, vertically, duties, force, vehicle, ingredient, cartridges, certification, signature, whenever, meet, compatibility, loading, hour, side, fixed, machinery, entry, after, shipownersand, candidate, shielding, un, port_state, stowed, except, transport, permissible, temperature, smoke, voyages, ensure, moulded, replaced, gross, testing, competence, globallast, regulation
TANK (253)	pallet, inflated, iso, parties, strength, accessories, mf, litre, fly, medical, hatches, type, access, monetary, react, tare, manual, flexible, pbz, officers, placed, escape, so, mm, exemptions, accepted, vehicle, performing, unsafe, denunciation, iv, particulars, surfaces, refrigerant, examination, indicated, damage_stability, verify, alternatively, sub-systems, plan, exceeding, cushioning, if, when, meets, applies, passenger, equivalent, minutes, normally, doors, alloy, having, melting_point, bulk, fittings, ventilated, above, diameter, carriage, freeboards, inspection, contain, reception_facilities, refer_to, unified, wearing, circuit, pfso, vent, state, two-thirds, low-toxicity, fissile, space, gastight, alia, remanufactured, distress, stations, inner, competence, where, mcconnell, accessible, due, duties, cargoes, watertight_doors, trifluoride, slop, referred, complying, should, safety, attached, material, endorse, aerosols, permit, packages, months, automatically, welded, officer, consist_of, organizations, survival_craft, covered, prohibited, skids, cargo_holds, pigments, water-resistant, requirements, compatibility, designated, aqueous, embarkation, mercuric, vertical_zone, self-heating, packaged, maintained, acid, date, ship, globallast, address, balconies, flammability, wearer, multiplied, duct, detection, cradles, furnishings, this, primer, consignment, entry, thereto, equipment, notifications, respect, steel, ballast_tanks, flange, exceeded, packed, word, arrangements, pursuant, mandatory, cartridges, accumulator, navigational, symbols, contact, measures, sanitary, interpretations, adopted, instructions, individually, consignor, consigned, machinery_space, purpose, situated, dispersible, checking, radio, measured, whenever, omb, forming, revised, specific, haop, located, res, explosive, relevant, applicator, muster, extent, article, liquids, certificates, valves, blank_flange, periodic, instrument, openings, shell_plating, contents, freeboard, side-wall, lifejackets, listed, containing, carbon, unit, garbage, seafarer, other, extinguishing, reconstituted, ballast_water, which, interim, fire_protection, nitric, paragraph, compatible, stopping, certified, equal, applicable, stairways, modified, combination, residue, pyrotechnic, flammable, vertically, cooling, bulk_cargoes, unless, materials, automatic, subparagraph, invasive, valve, relief, drums, subdivided, fire_extinguishing, assumed, instruction, distance, recognized, limits, solid, list, expelling, code, gross, regulation, person, emergency

BALLAST_WATER (195)	ventilation, issued, applicable, materials, than, weather_deck, foam, annex, prescribed, test, bilges, other, after, contents, appropriate, remanufactured, certification, equipment, accidental, type, walls, liquid, person, authorized, applicator, tanks, g, cargo, freight_container, unit, door, radionuclide, hydrogen, concerned, readily, subdivided, chemical, engaged, laid, packed, subjected, sill, identification, classified, fusible, below, attached, radio, flammable_solids, plastic, flexible, leakproof, transverse, cartridges, tubes, vehicle, non-refrigerated, freeboard, text, label, fire_pump, bar, surfaces, two-thirds, edp, permit, sulphide, accordance, violation, fire_alarm, documentation, officers, incineration, for, date, multiplied, sadt, disembarkation, protectors, xa, safe, seemp, radioactive_material, crude_oil_washing, unified, doors, escape, machinery, loading, radioactive, wetted, litre, self-reactive, port_facilities, non-conformity, boiler, least, vent, cylinder, outside, re-use, form, such, testing, bulkheads, of, vents, carriage, sealed, adhesive, less, the, manually, party, melting_point, arrangements, ducts, supplemented, special, seagoing, examination, refer_to, organization, witness, source, noise, sound, required, standard, displayed, requirements, account, miscellaneous, fly, approval, paragraph, designated, suitably, sts, non-fixed, deck, certified, space, freeboards, embarkation, provided, relief, declaration, provisions, whenever, segregation, flammable, manoeuvre, present, thermal, regulation, any, list, constructed, deformation, ecx, inert_gas, notified, convention, not, refrigerating, gross, chief_engineer, approving, entitled, un, carry, means, accession, may, metal, shall, shielding, used, measured, fishmeal, tested, ship, facility, by, life-saving, ensure, harmful_substances, permissible, be, enclosed, definitions, or, renewal, seafarer
BULK_CARGOES (71)	furnishings, pollutant, chemical, annex, chapter, below, shipownersand, conform, peroxide, code, maintained, periodical, sufficient, paragraph, shall, or, condition, competent, cargo, replaced, cargo_units, consigned, assigned, harmful_substances, any, appropriate, state, languages, may, chemically, msc, residue, fibre, transported, aqueous, hazards, ducts, revalidation, un, shielding, refer_to, containment, reference, resolution, acute, filled, compliance, hydrogen, less, containing, material, of, openings, bar, performing, entries, means, information, permit, tested, facility, subparagraph, constructed, liquefied_gases, by, valves, than, provisions, dangerous_goods, be, renewal
BULKHEAD_DECK (40)	be, spaces, requirements, may, cargo, valid, lightest, described, interpretations, harmful_substances, alarm, edp, corrosive, ventilated, iso, surveyed, ships, unless, cm, system, than, limits, verification, purpose, a-ii, adopted, administrations, facility, stowage, means, tanks, used, testing, medical, listed, alternative, shall, renewal, shielding, bulkhead
DANGEROUS_GOOD (12)	used, verifying, shall, necessary, ducts, each, may, harmful_substances, tested, duties, code, signature
EXPIRY_DATE (12)	fly, certified, shall, radio, instructions, security, denunciation, entitled, openings, specified, indicated, than
FIRE_ALARM (56)	means, msc, approval, be, by, provided, surveyed, noxious_liquid_substances, with, tested, certified, defined, depositary, closed, attachments, vapour, renewal, bulkhead, instruction, moira, placarding, width, aqueous, corrosive, oil_record_book, supplementary, protective, special, least, duct, transported, indicated, resolution, entry, contents, present, any, unit, ducts, provisions, procedures, list, recommended, shall, requirements, ibc, used, required, liquid, ensure, detection, permissible, ecx, unless, regulation, lightest
FREIGHT_CONTAINERS (46)	attachments, cargo_holds, near-coastal, or, compatibility, stations, accordance, openings, audit, capable, provisions, stowage, cargo_units, suitable, by, freeboards, alternatively, notifications, amended, state, harmful, least, equivalent, inner, defendant, machinery_spaces, inerting, particulars, than, machinery, composite, voyage, enters, tested, http, criteria, material, liquid, shall, accommodation_spaces, facility, issued, insulation, valve, ratification, fibreboard

GENERAL_PURPOSE _CONTAINER (4)	consideration, meets, sealed, may
GROSS_TONNAGE (108)	of, unless, liability, removable, ships, apparatus, baggage, other, arrangements, inserted, comply_with, sealed, kg, applicable, ship's, by, particles, lightest, ibc, certificate, combustion, shipownersand, code, hexafluoride, closed, paragraph, cylinder, be, ro-ro, vapour, fuze, inspected, accordance, transboundary, provisions, information, containers, carrying, rhumb_line, administration, may, package, declaration, fitted, emergency, inert_gas, competence, chemical, satisfied, under, satisfaction, necessary, openings, appropriate, not, msc, organization, consignor, any, complies, kept, means, valid, asbestos, except, into, testing, protected, fittings, certifying, signature, res, certificates, listed, multiwall, trimming, facility, distress, volumetric, cargo_holds, harmful, plastics, safe, flashpoint, harmful_substances, ducts, required, recommended, approving, amending, regulation, cargo, tested, rigid, subdivision, provided, renewal, shall, shielding, dangerous, residue, subjected, least, or, ship, than, ensure, permissible
INTERNATIONAL_ CONVENTION_ ON_LOAD_LINES (11)	forward_perpendicular, accept, way, binary, obtain, idcm, emerged, immediately, iridium, developer, SOLAS
LIQUEFIED_GASES (41)	such, ship, tanks, fibreboard, fracture, proper, surveyed, performing, fissile, seafarer, integrity, recommended, bar, notify, ducts, information, issued, flammability, accession, than, mm, stowed, denunciation, facility, edp, listed, be, shall, tested, of, filling, liquid, provisions, portable, tank, by, ships, pressure, manufacturer's, liquefied, renewal
MACHINERY_SPACES (113)	lifejackets, proficiency, made, protection, limitations, possessing, fire-fighting, paragraph, gaseous, slop, including, emergency_escape, test, of, accessible, normal_conditions, waste, plywood, imdg_code, portable, msc, tank, pressure, ship, ensure, identification, requirements, deposited, substances, oily, bottom, handling, defined, documented, indicated, nothing, protective, near-coastal, residue, ballasting, stability, emergency, openings, necessary, establish, inflatable, shell_plating, dangerous, into, adequate, lithium, verification, gross, inflated, protected, peroxide, months, shell, ibc, cylinders, unless, navigational, facility, load, secretary-general, alternative, officer, undertake, otherwise, transported, temperature, before, pyrotechnic, fire_pumps, power-driven, situated, may, inspection, applicator, harmful, unpackaged, provided, languages, fusible, seafarer, ducts, edp, equipment, permissible, cartridge, shielding, bulkhead, flag, ships, doors, competence, used, bearing, instruction, not, less, or, renewal, least, adopted, cargo, noxious_liquid_substances, shall, suitable, harmful_substances, machinery_space, spaces, be
MEDICAL_FIRST_ AID_GUIDE (12)	precautions, manhole, substances, msc, sealed, facility, defined, satisfaction, ensure, permissible, be, by
NOXIOUS_LIQUID _ SUBSTANCES (33)	openings, lockers, ship, co-operate, package, any, engaged, information, regulation, supplied, ducts, residue, surveyed, escape, such, hazard, category, amendments, gross, means, noxious_liquid_substances, substance, lightest, subparagraph, or, flammable, tested, certify, unless, substances, cargo, filling, certificate
OIL_RECORD_BOOK (18)	flammable, shielding, such, label, comply_with, minimum, isolation, transmitting, ducts, removable, lightest, permissible, compatibility, embarkation, after, purposes, ship, shall

OIL_TANKERS (73)	of, unattended, requirements, cargo, ship, pressure, uptake, shall, pumps, accordance, steel, satisfaction, ?, electrical, inserted, means, carrying, not, breathing, damage, vent, special, a-iii, indicated, expelling, parties, authorized, plan, fire_extinguishers, iii, surveyed, require, accession, used, measured, lightest, ducts, ecx, below, auxiliary_machinery, ensure, less, manhole, permissible, removable, heat, shielding, ui, titanium, filled, information, fire-extinguishing, maintained, reasonably, hazard, suitable, permanent, pollutant, be, codes, vacuum, facility, renewal, than, or, regulation, accidental, tested, testing, survival_craft, by, tissue, harmful_substances
ON_BOARD (251)	fire_pump, issued, solas, filling, other, mg, withdrawn, vent, disembarkation, solid, pumping, assigned, loading, flexible, ignition, loads, verification, hazard, pesticides, un, durably, noxious_liquid_substances, certification, account, flag, labelled, ambient, acceptance, inspection, removable, embarkation, as, uptake, ventilation, ibcs, package, expiry_date, authentic, maintained, sidescuttles, resolution, parties, necessary, atmospheric_pressure, containment, recommended, temperature, sediments, intended, periodic, adversely, excess, immersion, taking, unloading, ui, g, contents, space, consigned, measured, nobob, cargo_holds, distress, compatible, rigid, applicator, columns, remanufactured, rhumb_line, transverse_bulkhead, megc, compatibility, standard, sulphide, obstructions, training, category, indicated, provided, seafarer, mpgm, attached, paragraph, inserted, accordance, drum, validity, competent, annex, lighting, detection, properly, liquefied, granular, test, protected, kg, stations, plies, pigments, breathing_apparatus, notification, nozzles, correspond, into, equally, ships, shipownersand, deposit, potassium, overturning, mandatory, discharges, segregation, amended, compliance, oil_record_book, asbestos, two-thirds, contain, inter, forth, fibre, organization, samples, substances, treated, materials, plastics, unless, of, containing, than, flammable, mass, appropriate, tank, specimen, satisfy, cartridge, periodical, column, stairways, width, wastes, piping, regard, amendments, haop, edp, documentation, suction, survival_craft, venting, gross, suitably, length, radioactive_material, relief, satisfaction, officers, machinery_space, soon, a-ii, fitted, labelling, list, person, gases, declaration, re-use, baggage, defined, stowage, nothing, alternatively, res, determine, cargo_units, self-reactive, bulkheads, convention, recognized, whenever, fly, bulk_cargoes, adjacent_spaces, code, cargo, lifting, may, means, power-driven, combustion, phosphorus, cylinder, special, lifejackets, ensure, used, prevent, under, such, charge, master, damage, regulation, entitled, transmitting, chemical, aqueous, msc, globallast, administration, port_facilities, refer_to, automatically, any, requirements, shielding, harmful_substances, iaea, unpackaged, except, certified, for, liquefied_gases, certifying, required, renewal, by, equipment, least, sound, certificate, containers, manifold, request, lightest, ducts, aluminium, peroxides, not, permissible, facility, shall, ship, or, tested, be
RESCUE_BOATS (39)	notify, below, ship, periodic, fracture, whenever, prevent, sufficient, paragraph, materials, requirements, class, omb, nominated, edp, attachments, present, hazard, distress, unpackaged, facility, copies, containers, integrity, specified, than, provided, certificate, relief, shielding, ducts, by, into, renewal, medical_first_aid_guide, satisfied, be, harmful_substances, shall
STEERING_GEAR (47)	or, cargo, any, automatic, requirements, whenever, detection, ventilated, necessary, concerned, linings, avoid, rudder_stock, transmit, subdivided, lightest, ducts, packed, hazard, with, fly, least, un, auxiliary_machinery, accommodation, less, absorb, unpackaged, space, comply_with, cushioning, chapter, shielding, iso, seafarer, provided, combination, jerricans, required, kg, ship, harmful_substances, noxious_liquid_substances, be, renewal, facility, shall
SURVIVAL_CRAFT (81)	amending, such, dust, specify, draught, watchkeeping, flashpoint, tested, include, msc, neutron, gross, used, capable, nitrates, organization, cargo, requirements, regulation, accordance, radioactive, provided, present, freeboard_deck, means, flame, lithium, special, column, anti-fouling, fitted, ship, pump-room, flammable, cease, volumetric, safe, under, ducts, procedures, referred, chapter, pyrotechnic, impaired, control_stations, should, oil_record_book, deposit, necessary, drills, standard, of, render, haop, training, maintained, accession, arrangements, load, manufactured, except, by, or, certification, wetted, strength, intended, properly, whenever, ensure, flash, communicated, within, least, spaces, harmful_substances, permissible, be, renewal, shall, facility

Appendix E: Collocates for Collocation Networks

Source	Target	Score	Source	Target	Score
accordance	WITH	144.92	administration	SATISFACTION	80.99
accordance	IN	81.98	administration	ENTRUST	46.34
accordance	ARTICLE	41.88	administration	THE	40.85
accordance	PROVISIONS	33.31	administration	APPROVED	39.12
accordance	CALCULATED	27.29	administration	OPINION	36.13
accordance	PARAGRAPH	26.96	administration	BY	35.46
accordance	REGISTRATION	25.07	administration	RECOGNIZED	32.68
accordance	DEMONSTRATING	24.94	administration	MAY	31.20
accordance	PUBLICATION	22.29	administration	SATISFIED	29.90
accordance	REGULATION	21.16	administration	WAIVE	27.27
accordance	DETERMINED	21.14	administration	ACTING	25.23
accordance	LAW	17.49	administration	NOMINATING	21.05
accordance	COLOURED	17.09	administration	GRACE	21.00
accordance	BEEN	15.67	administration	SANCTION	20.44
accordance	STOWAGE_PLAN	15.49	administration	CONSIDERS	20.08
accordance	VALUATION	15.49	administration	PERMIT	19.93
accordance	SUBPARAGRAPH	15.07	administration	DISPENSE	19.37
accordance	TESTED	14.42	administration	EXTEND	18.30
accordance	ANDAPPROVED	13.99	administration	DEEMS	17.56
accordance	THIS	13.83	administration	ACCEPTABLE	17.03
accordance	FORCE	13.80	administration	REQUEST	16.77
accordance	PROCEDURE	12.52	administration	ASSUMES	16.51
accordance	METHOD	12.47	administration	ALLOWS	16.34
accordance	PROCEDURES	12.36	administration	CONCERNED	15.61
accordance	XIV	12.09	administration	BEHALF	15.22
accordance	DENOUNCE	11.75	administration	REQUESTING	15.18
accordance	CLASSIFIED	11.50	administration	ALLOW	15.11
accordance	MOREOVER	11.19	administration	EARLIEST	15.07
accordance	ESTABLISHED	11.15	administration	ENDORSE	14.92
accordance	VERSION	10.73	administration	FURNISH	14.15
accordance	PERFORMED	10.62	administration	NOTIFY	13.88
accordance	PARAGRAPHS	10.53	administration	PILGRIM	13.70
accordance	REGULATIONS	10.37	administration	SHIP_OWNER	13.70
accordance	HARMFUL_SUBSTANCE	10.31	administration	BELHALF	13.70
accordance	METHODS	10.25	administration	PROLONG	13.70
accordance	STANDARD	10.20	administration	DETERMINED	13.62
accordance	ENTER	9.86	administration	HOWEVER	13.20
accordance	XVI	9.79	administration	INFORM	12.84
accordance	QUALIFIED	9.78	administration	GUARANTEE	12.52
accordance	REQUIREMENTS	9.76	administration	OPPORTUNITY	12.52
accordance	CONDUCTED	9.63	administration	ORGANIZATION	12.48
accordance	EFFECT	9.53	administration	GOVERNMENT	12.47
accordance	FURNISH	9.41	administration	SURVEYOR	12.10
accordance	MANUFACTURER'S	9.29	administration	EXTENDED	11.95
accordance	SERIALLY	9.17	administration	OMIT	11.80
accordance	AUDITS	8.85	administration	WITHDRAWN	11.49
accordance	MARKED	8.81	administration	IMPRACTICABLE	11.23
accordance	A-VI	8.71	administration	PRESET	11.18
accordance	SECTION	8.60	administration	NEWLY-CONSTRUCTED	11.18
accordance	FUNDS	8.48	administration	DEVELOPED	11.16

Source	Target	Score	Source	Target	Score
amendment	ENTERS	63.53	approval	RATIFICATION	196.86
amendment	FORCE	61.94	approval	ACCEPTANCE	160.23
amendment	PROPOSED	50.00	approval	ACCESSION	143.22
amendment	AN	47.90	approval	PLATE	68.51
amendment	ENTER	46.20	approval	FORDESIGN	56.90
amendment	INTO	45.05	approval	DEPOSITED	53.47
amendment	DEEMED	40.35	approval	INSTRUMENT	50.82
amendment	CONFERENCE	38.61	approval	REQUISITE	44.42
amendment	REVISION	34.38	approval	SIGNATURE	42.52
amendment	BOUND	31.39	approval	FOLLOWED	41.94
amendment	CONCERNING	31.04	approval	MULTILATERAL	40.06
amendment	DECLINED	30.79	approval	MANUFACTURE'S	37.86
amendment	CIRCULATED	30.20	approval	DESIGN	36.68
amendment	PROPOSE	28.45	approval	CASE-BY-CASE	30.83
amendment	ANNEX	28.11	approval	EFFECTED	30.40
amendment	ADOPTION	26.57	approval	UNILATERAL	29.99
amendment	OBJECTED	26.36	approval	RETESTED	28.45
amendment	ACCEPTS	26.13	approval	COUNTRY	27.60
amendment	ACCEPTED	24.30	approval	OR	27.35
amendment	ACCEPT	23.89	approval	AA	26.62
amendment	OBJECTION	22.85	approval	SUBJECT	26.55
amendment	ENTRY	22.72	approval	PLATES	25.12
amendment	PRESENT	22.28	approval	SAFETY	24.08
amendment	SUBPARAGRAPH	21.51	approval	BODY	22.95
amendment	REJECTED	20.92	approval	COMPETENT	21.02
amendment	PROTOCOL	20.85	approval	AUTHORITY	20.39
amendment	DECIDES	20.39	approval	FOLLOWEDBY	18.93
amendment	DENOUNCE	20.01	approval	FIRST_LINE	18.93
amendment	ARTICLE	18.72	approval	INSTRUMENTS	18.23
amendment	OBJECT	17.76	approval	SHIPMENT	18.22
amendment	UNANIMOUS	17.41	approval	TYPE	17.25
amendment	DELETION	17.41	approval	WATER-BASED	16.80
amendment	CONVENTION	17.18	approval	NUMBER	16.66
amendment	WHICHEVER	17.16	approval	DECLARE	15.71
amendment	CONSIDERATION	16.93	approval	SERIAL_NUMBER	14.70
amendment	ACCEPTANCE	16.64	approval	TACIT	14.07
amendment	FORWARDING	16.46	approval	MAXIMUM_GROSS_WEIGHT	14.07
amendment	THEREUPON	16.46	approval	ACEP	13.65
amendment	TO	15.94	approval	MANUFACTURE	13.49
amendment	ADOPTED	15.74	approval	AUSTRIAN	13.31
amendment	DATE	14.83	approval	VESTED	13.31
amendment	HAVE	14.77	approval	ORACCESSION	13.31
amendment	ANY	14.70	approval	TESTING	13.21
amendment	PROCEDURES	14.35	approval	DESIGNS	12.64
amendment	FORMALLY	14.15	approval	EXPLICIT	12.53
amendment	SUB-CHAPTER	14.15	approval	LETTERS	12.50
amendment	APPLICABLE	14.00	approval	CONFORMITY	11.93
amendment	BECAME	13.39	approval	ACCOMPANY	11.71
amendment	BLUE	13.39	approval	REQUIRE	11.60
amendment	EXTENSIVELY	13.39	approval	CERTIFICATE	11.30

Source	Target	Score	Source	Target	Score
code	STCW	86.63	competent	AUTHORITY	244.47
code	IMDG	64.46	competent	COUNTRY_OF_ORIGIN	50.67
code	ISPS	59.78	competent	COUNTRY	43.40
code	THIS	57.20	competent	AUTHORITIES	41.40
code	IMSBC	47.16	competent	IES	35.08
code	HSC	46.51	competent	BY	31.92
code	PART	45.01	competent	APPROVED	31.89
code	PROVISIONS	38.88	competent	BODY	31.38
code	SUBJECT_TO	35.63	competent	AUTHORIZED	27.24
code	PRESSURE-VESSEL	35.36	competent	PORT_OF_LOADING	23.31
code	SEAFARERSTRAINING	32.23	competent	ROAD	22.93
code	ISM	29.92	competent	THE	21.06
code	INTERNATIONAL	29.75	competent	APPROVAL	21.02
code	PRACTICE	28.52	competent	COURT	20.54
code	FIRE	22.70	competent	SKILLED	17.83
code	FSS	22.29	competent	APPROVING	16.60
code	IMPLEMENTATION	21.92	competent	RECOGNIZED	15.65
code	SAFETY	20.51	competent	AUSTRIA	15.46
code	THE	20.15	competent	SELECTIVE	15.46
code	OF	19.71	competent	PUBLICAUTHORITY	14.56
code	FTP	18.61	competent	SENT	13.24
code	ALPHANUMERIC	17.36	competent	QUALITY-ASSURANCE	12.84
code	BLU	17.23	competent	ACCOMPANY	12.84
code	CHEMICAL	16.88	competent	NOTIFY	11.58
code	ADOPTED	16.74	competent	SATISFIES	11.50
code	INF	16.58	competent	QUALITY_ASSURANCE	11.18
code	B-I	16.32	competent	ORGANIZATIONS	11.17
code	DESIGNATING	15.85	competent	SATISFACTION	11.11
code	INTERNATIONAL_MARITIME _DANGEROUS_GOODS_CODE	15.73	competent	UNLESS	11.09
code	A-I	15.38	competent	APPROVALS	11.06
code	VRI	15.02	competent	UNILATERAL	10.75
code	ETHICS	15.02	competent	DEEMED	10.36
code	MARITIME	14.44	competent	AUTHOR	10.20
code	EXECUTION	14.20	competent	REPRODUCIBLE	10.20
code	HIGH-SPEED	14.18	competent	BK	10.20
code	RECOMMENDATORY	13.65	competent	INFRINGEMENTS	10.20
code	AMENDED	13.63	competent	DETERMINED	9.84
code	BULK	13.47	competent	TRAINED	9.62
code	SECTION	13.37	competent	ACCEPTABLE	9.58
code	WATCHKEEPING	13.21	competent	CONSULTING	9.55
code	SIONS	13.18	competent	QUALITY_CONTROL	9.55
code	MEANS	12.97	competent	ACCREDITED	9.55
code	CERTIFICATION	12.79	competent	CHANNEL	9.39
code	NOX	12.56	competent	PERSON	9.31
code	CHAPTER	12.52	competent	ILLUSTRATION	9.08
code	SUBSECTIONS	12.52	competent	COUNTRIES	8.95
code	PREAMBLE	12.42	competent	CONCERNED	8.75
code	A-III	12.29	competent	EXPERT	8.67
code	COMPARE	12.28	competent	AUDITED	8.67
code	LEGALLY	12.25	competent	IDENTITY	8.65

Source	Target	Score	Source	Target	Score
convention	THIS	73.56	may	BE	61.12
convention	PRESENT	67.57	may	ADMINISTRATION	31.10
convention	INTERNATIONAL	42.67	may	LIQUEFY	27.15
convention	LIABILITY	38.82	may	BECOME	24.62
convention	THE	35.98	may	DENOUNCED	24.23
convention	FORCE	35.18	may	USED	21.24
convention	SCHEDULE	33.13	may	WAIVED	20.30
convention	PARTIES	27.85	may	WAIVE	18.86
convention	BASEL	27.65	may	WHICH	18.21
convention	OPRC	27.16	may	INCLUDE	17.85
convention	FUND	26.74	may	PERMIT	17.62
convention	NOTHING	25.84	may	PARTY	17.08
convention	VIII	25.03	may	ENTRUST	16.81
convention	AMENDED	24.24	may	LONGER	16.59
convention	SOLAS	23.41	may	TRANSPORTED	16.45
convention	PREVENTION_OF_POLLUTION	23.06	may	VARY	16.33
convention	ARTICLE	23.04	may	REDUCED	15.94
convention	REMOVAL	22.36	may	ACCEPT	15.07
convention	WRECK	22.11	may	ALSO	14.77
convention	TO	21.93	may	INCREASED	14.76
convention	AMENDMENTS	21.89	may	MODIFY	14.74
convention	STCW	21.77	may	EXEMPT	14.37
convention	ACCEDED	21.65	may	PROPOSE	14.32
convention	ATHENS	21.61	may	CHAIN_LOCKER	14.19
convention	APPLY_TO	20.22	may	ARISE	13.71
convention	CIVIL	20.07	may	ASSISTANCE	13.52
convention	INTO	19.89	may	ACCEPTED	13.48
convention	ANNEX	19.61	may	ALLOW	13.35
convention	RATIFIED	19.49	may	SUCH	12.79
convention	SIGNED	18.91	may	UNCHANGED	12.76
convention	PROTOCOL	18.67	may	SUBMITTING	12.76
convention	SIGNATURE	18.59	may	UNPACKAGED	12.69
convention	SEAFARERSRIGHTS	18.19	may	DETAIN	12.52
convention	HNS	17.80	may	REQUIRE	12.19
convention	PARTY	17.80	may	INSPECT	11.79
convention	KONG	17.56	may	AMENDED	11.75
convention	HONG	17.56	may	REQUEST	11.64
convention	AMENDMENT	16.82	may	WITHDRAW	11.63
convention	CSC	16.79	may	THIS	11.48
convention	OF	16.77	may	SOME	11.44
convention	SUCCESSOR	16.73	may	RELAXATIONS	11.42
convention	ENTRY	16.72	may	DISPENSE	11.42
convention	APPLY	16.67	may	BARS	11.41
convention	DIVERSITY	16.46	may	BY	11.38
convention	NATIONS	16.42	may	AUTHORIZE	11.34
convention	ACCESSION	16.42	may	INVOKE	11.32
convention	UNDER	16.18	may	VALIDITY	11.26
convention	ESTABLISHMENT	15.68	may	STATE	11.20
convention	AMENDING	15.57	may	CONTINUE	11.19
convention	STATES	15.54	may	EXTEND	11.17

Source	Target	Score	Source	Target	Score
organization	DEVELOPED	88.39	packagings	INNER	151.48
organization	BY	69.66	packagings	OUTER	110.47
organization	ADOPTED	65.72	packagings	LARGE	90.20
organization	RESOLUTION	63.60	packagings	INTERMEDIATE	79.46
organization	GUIDELINES	60.35	packagings	BAGS	69.88
organization	SECRETARY-GENERAL	51.27	packagings	COMBINATION	64.37
organization	RECOGNIZED	48.99	packagings	MET	48.61
organization	THE	45.43	packagings	PLASTICS	40.78
organization	CONSULTATIVE	43.50	packagings	IBCS	38.71
organization	SURVEYOR	41.76	packagings	PACKING_INSTRUCTION	37.68
organization	CONVENE	41.28	packagings	HERMETICALLY	35.48
organization	CONVENED	35.37	packagings	COMPOSITE	35.27
organization	REFER_TO	33.18	packagings	PAPER	29.93
organization	MARITIME	30.60	packagings	AUTHORIZED	29.14
organization	MEMBERS	29.20	packagings	POLYGONAL	28.85
organization	INSTITUTION	28.56	packagings	FIBREBOARD	28.43
organization	COMMUNICATE	27.81	packagings	ARE	27.72
organization	INTER-GOVERNMENTAL	26.55	packagings	DRUMS	27.22
organization	HEADQUARTERS	26.50	packagings	METAL	26.92
organization	NOMINATED	25.99	packagings	PACKED	25.88
organization	INTEGRATION	24.23	packagings	NET	24.21
organization	RECOMMENDATIONS	23.66	packagings	STRONG	23.48
organization	COMMITTEE	23.48	packagings	RECONDITIONED	22.98
organization	NOTIFY	22.73	packagings	LEAD-FREE	22.70
organization	STANDARDIZATION	22.26	packagings	RECEPTACLES	21.74
organization	INTERNATIONAL	21.24	packagings	PAIL-SHAPED	20.40
organization	INFORM	20.48	packagings	POLYVINYLIDENE	20.40
organization	PARTIES	19.94	packagings	WATER-SOLUBLE	20.40
organization	AUDIT	17.95	packagings	TAPER-NECKED	20.40
organization	CIRCULATE	17.87	packagings	WATERPROOFED	20.40
organization	HEREINAFTER	17.04	packagings	SEALED	19.97
organization	ORGANS	16.75	packagings	CONFORM	19.12
organization	DISSEMINATE	16.75	packagings	GENERAL	18.73
organization	AMENDMENTS	16.08	packagings	BOXES	18.60
organization	NON-GOVERNMENTAL	15.98	packagings	SALVAGE	18.50
organization	COLLATE	15.76	packagings	GLASS	17.98
organization	CIRCULATION	15.47	packagings	UN	17.95
organization	ORGANIZATIONS	15.09	packagings	CONSISTING	17.91
organization	PARTICIPATE	14.50	packagings	FOLLOWING	17.72
organization	SCHEME	14.43	packagings	PORCELAIN	16.79
organization	ECONOMIC	14.07	packagings	DISTRIBUTORS	16.50
organization	WORLD	13.82	packagings	RE-USED	16.07
organization	REGIONAL	13.82	packagings	UNCLEANED	15.90
organization	VIRTURE	13.72	packagings	TEXTILE	15.80
organization	FAILINGS	13.72	packagings	JERRICANS	15.78
organization	LOSING	13.57	packagings	NAILED	15.18
organization	RECEIVES	13.29	packagings	DIVIDERS	15.18
organization	LABOUR	13.14	packagings	KRAFT	14.98
organization	MATTERS	13.08	packagings	MASS	14.91
organization	PERSON	12.98	packagings	DIFFER	14.87

Source	Target	Score	Source	Target	Score
provided	EXPRESSLY	60.57	provisions	SUBJECT_TO	53.62
provided	OTHERWISE	41.07	provisions	GENERAL	43.07
provided	BE	31.67	provisions	CODE	38.85
provided	THAT	28.48	provisions	THIS	37.76
provided	UNLESS	25.75	provisions	SPECIAL	34.83
provided	SHALL	22.75	provisions	ACCORDANCE	33.29
provided	EXCEPT	21.58	provisions	OF	32.50
provided	ARE	20.19	provisions	CHAPTER	28.65
provided	MEANS	19.24	provisions	PACKING	27.81
provided	WITH	17.43	provisions	THE	27.17
provided	UNCHANGED	17.36	provisions	VIII	26.70
provided	GENERAL	15.79	provisions	NOTWITHSTANDING	25.09
provided	THEY	15.29	provisions	COMPLY_WITH	23.07
provided	AUTHORIZED	15.15	provisions	FOREGOING	22.54
provided	COMPLETES	14.24	provisions	RELEVANT	21.38
provided	OAR	13.49	provisions	MET	21.22
provided	STOPPING	11.98	provisions	CONFORM	19.82
provided	ARRANGEMENTS	11.69	provisions	WITH	19.62
provided	PIPING_ARRANGEMENT	11.21	provisions	APPLICABLE	19.45
provided	RELIEF-DEVICE	10.92	provisions	ARE	18.66
provided	CLOSED-OFF	10.92	provisions	ARTICLE	18.03
provided	AMENDMENTS	10.71	provisions	SEGREGATION	16.49
provided	VENTILATION	10.51	provisions	APPLY	16.05
provided	FOLLOWING	10.30	provisions	COMPLYING	15.98
provided	PROBABILITIES	10.17	provisions	APPLY_TO	15.70
provided	FOAM-GENERATING	10.17	provisions	SPECIFIC	15.59
provided	PARAGRAPH	9.77	provisions	CONVENTION	15.35
provided	SEGREGATED_BALLAST	9.75	provisions	SECTION	14.84
provided	INFORMATION	9.67	provisions	MEET	14.68
provided	BOOKLET	9.41	provisions	TRANSITIONAL	14.31
provided	THOLE	9.38	provisions	AUTHORIZED	14.30
provided	COORDINATES	9.36	provisions	RECOMMENDATORY	14.11
provided	MESS_ROOM	9.36	provisions	REGULATION	14.05
provided	PACKAGINGS	9.12	provisions	INTRODUCTORY	13.21
provided	VACUUM_VALVE	8.92	provisions	EFFECT	13.14
provided	CREDENTIALS	8.92	provisions	SUPPLEMENT	13.05
provided	TWENTY-ONE	8.92	provisions	MANDATORY	12.60
provided	STRETCHING	8.92	provisions	CINNABAR	12.27
provided	UNMANAGED	8.92	provisions	STOWAGE	11.76
provided	FUNDS-IN-TRUST	8.92	provisions	ADDITIONAL	11.14
provided	TIMER	8.92	provisions	SUBSECTION	11.11
provided	ORIGINATED	8.92	provisions	CONTAINS	11.09
provided	RELEASABLE	8.92	provisions	PART	10.80
provided	LOW-STRESS	8.92	provisions	COMPLIED	10.27
provided	AUDIO-VISUAL	8.92	provisions	PARAGRAPH	10.04
provided	FACILITIES	8.59	provisions	IBCS	9.98
provided	LONGITUDINAL	8.50	provisions	MEETS	9.91
provided	DENUNCIATIONS	8.32	provisions	SEED	9.75
provided	VENEERS	8.32	provisions	ENFORCEMENT	9.68
provided	EMERGENCY_STEERING	8.22	provisions	EXTENDED	9.55

Source	Target	Score	Source	Target	Score
regulation	THIS	61.43	required	BY	36.56
regulation	PARAGRAPH	42.55	required	REGULATION	29.62
regulation	ANNEX	39.76	required	COMPETENCE	26.79
regulation	SOLAS	34.79	required	EVIDENCE	24.28
regulation	REQUIRED	29.61	required	DEMONSTRATE	23.43
regulation	COMPLY_WITH	27.42	required	SEEMP	21.45
regulation	PURPOSE	24.97	required	PARAGRAPH	20.43
regulation	JUNE	23.86	required	JETS	19.50
regulation	IN	22.55	required	EEDI	16.50
regulation	ACCORDANCE	21.16	required	FIRE_PUMPS	16.24
regulation	REQUIREMENTS	19.62	required	PROFICIENCY	15.47
regulation	AS	18.94	required	WEATHERING	15.10
regulation	NOTHING	18.64	required	AS	14.78
regulation	REFERRED	18.63	required	INFORMATION	14.75
regulation	APPLIES	18.60	required	ACHIEVED	14.57
regulation	MARPOL	17.44	required	RETEST	14.14
regulation	OIL_TANKERS	17.19	required	IS	13.82
regulation	CONVERSIONAS	16.77	required	MARKINGS	13.68
regulation	UNDER	16.38	required	SIMULTANEOUS	13.26
regulation	EC	15.89	required	RADIO_INSTALLATION	12.29
regulation	OF	15.42	required	IECC	12.25
regulation	AJOR	15.31	required	SELF-ACTIVATING	12.25
regulation	WITH	14.29	required	RELIEF-DEVICE	12.25
regulation	PROVISIONS	14.06	required	CANDIDATE	12.23
regulation	RENUMBERING	13.83	required	CERTIFICATION	12.22
regulation	VERTICAL_ZONING	13.83	required	UNDERSTANDING	11.81
regulation	COMPLYING	13.62	required	FIRE_ALARM	11.67
regulation	APPLY_TO	12.29	required	STANDARD	11.43
regulation	PURSUANT	12.10	required	ULTRAVIOLET	10.97
regulation	IECC	11.86	required	EPIRB	10.97
regulation	DECEMBER	11.22	required	MF	10.97
regulation	INSTALLING	11.22	required	HAVING	10.82
regulation	VHF_RADIO	11.22	required	UNDERTAKE	10.80
regulation	EPIRB	10.60	required	PROVIDE	10.73
regulation	INTERMEDIATE_SURVEY	10.02	required	PASS	10.61
regulation	SEEMP	9.86	required	OCCASIONALLY	10.53
regulation	MCGEE	9.68	required	TO	10.46
regulation	EAVY	9.68	required	LABEL	10.29
regulation	SARAH	9.68	required	CHAPTER	10.02
regulation	INSECURE	9.68	required	OVERFLOW-CONTROL	10.00
regulation	SINGLE-HULL	9.68	required	PURITY	10.00
regulation	BUOYANT_LIFELINES	9.68	required	RROSIVESUBSIDIARY	10.00
regulation	SPECIFIED	9.54	required	POINT_OF_DEPARTURE	10.00
regulation	TANKERS	9.33	required	BUOYANT_LIFELINES	10.00
regulation	SATELLITE	9.22	required	COLLEAGUES	10.00
regulation	EXCEEDED	9.10	required	NAVIGATION_LIGHTS	10.00
regulation	DEADWEIGHT	8.82	required	FROMIS	10.00
regulation	BULK_CARRIER	8.65	required	EQUATION	9.93
regulation	MARINE_CASUALTIES	8.65	required	ADDITION	9.75
regulation	METHODOLOGY	8.37	required	SURVEYS	9.64

Source	Target	Score	Source	Target	Score
requirements	COMPLY_WITH	63.36	shall	BE	151.87
requirements	COMPLYING	48.41	shall	SECRETARY-GENERAL	22.56
requirements	MEET	41.09	shall	EACH	22.35
requirements	FUNCTIONAL	38.00	shall	PROVIDED	22.13
requirements	COMPLIES	32.99	shall	NOT	21.08
requirements	MANDATORY	29.84	shall	SUCH	19.67
requirements	SATISFYING	28.72	shall	ENSURE	19.34
requirements	DAMAGE_STABILITY	24.51	shall	COMPLY_WITH	19.13
requirements	SEAFARERSRIGHTS	23.34	shall	INFORM	18.63
requirements	MEETING	23.11	shall	TAKE	18.35
requirements	CHAPTER	22.67	shall	NOTHING	18.23
requirements	REPORTING	18.32	shall	ENTER	18.10
requirements	COMPLIANCE	18.24	shall	INCLUDE	17.67
requirements	SECTION	18.01	shall	APPLY_TO	17.49
requirements	APPLICABLE	17.86	shall	KEPT	17.25
requirements	THIS	17.23	shall	ARRANGED	16.99
requirements	ADDITION	17.14	shall	EXCEED	16.74
requirements	REGULATION	17.10	shall	FITTED_WITH	16.48
requirements	WITH	16.64	shall	SO	16.38
requirements	PARAGRAPH	16.57	shall	DESIGNED	16.02
requirements	PRESCRIPTIVE	14.90	shall	MADE	15.81
requirements	SAFE_WORKING_LOADS	14.14	shall	EFFECTED	15.66
requirements	OF	13.82	shall	MARKED	15.57
requirements	MEETS	13.77	shall	HAVE	15.45
requirements	SATISFY	13.69	shall	VALID	15.41
requirements	THE	13.67	shall	STOWED	15.33
requirements	REGULATIONS	13.55	shall	APPLY	14.97
requirements	GENERAL	13.44	shall	FITTED	14.89
requirements	RELEVANT	13.15	shall	NOTIFY	14.75
requirements	NOTWITHSTANDING	13.10	shall	DEEMED	14.64
requirements	CONFORMITY	13.10	shall	CAPABLE	14.58
requirements	FOR	12.98	shall	CERTIFICATE	13.78
requirements	WAIVE	12.95	shall	PROTECTED	13.48
requirements	CERTIFICATION	12.95	shall	PARTY	13.41
requirements	SBT	12.95	shall	USED	13.24
requirements	DOCUMENTARY	12.49	shall	DEPOSITARY	13.09
requirements	FULFIL	12.34	shall	THIS	13.03
requirements	OPERATIONAL	12.31	shall	SECURED	12.43
requirements	FORMALITIES	151.87	shall	TREATED	12.19
requirements	RELAXATIONS	22.56	shall	RELIEF	11.79
requirements	AUDIBILITY	22.35	shall	DEPOSITED	11.73
requirements	ELECTRO-STATIC	22.13	shall	LEAST	11.62
requirements	MET	21.08	shall	FORCE	11.61
requirements	OMIT	19.67	shall	REPORT	11.58
requirements	PARTIALLY_ENCLOSED_LIFEBOATS	19.34	shall	CONFORM	11.41
requirements	HAND_FLARES	19.13	shall	CONSIST_OF	11.40
requirements	PORT_STATE_CONTROL	18.63	shall	CARRIED	11.33
requirements	PARAGRAPHS	18.35	shall	EVERY	11.33
requirements	ADDITIONAL	18.23	shall	ACCESSION	11.32
requirements	GOVERN	18.10	shall	DEVICES	11.01

Source	Target	Score	Source	Target	Score
spaces	RO-RO	91.88	specified	IN	44.09
spaces	ACCOMMODATION	68.17	specified	PERIOD	29.70
spaces	CATEGORY	65.53	specified	LONGER	25.74
spaces	CONTROL_STATIONS	64.92	specified	PARAGRAPH	23.36
spaces	VOID	54.99	specified	COMPETENCE	23.03
spaces	ENCLOSED	52.90	specified	SECTION	21.64
spaces	MACHINERY	48.84	specified	UNALLOYED	18.82
spaces	TRUNKS	47.85	specified	LIMITS	18.63
spaces	CARGO	44.07	specified	MEET	17.27
spaces	VEHICLE	43.08	specified	SUBPARAGRAPHS	17.12
spaces	GALLEYS	42.69	specified	TESTS	16.71
spaces	DOUBLE-BOTTOM	37.21	specified	AS	16.49
spaces	SERVICE	35.51	specified	TABLE	16.37
spaces	DOUBLE-HULL	32.68	specified	TIMEFRAME	16.30
spaces	SPECIAL	31.77	specified	SUBJECTED	16.09
spaces	ADJACENT	25.48	specified	PARAGRAPHS	15.97
spaces	STATIONS	24.94	specified	BELOW	15.28
spaces	AUTO-TELEPHONE	24.30	specified	AUSTENITIC	15.25
spaces	REEFER	23.85	specified	STEELS	15.23
spaces	OPEN	23.47	specified	STANDARD	15.03
spaces	MACHINERY_SPACES	23.17	specified	PERIODS	14.61
spaces	AIR-CONDITIONING	22.65	specified	INSOLATION	14.11
spaces	CONTAINING	21.80	specified	VALUES	13.98
spaces	CORRIDORS	21.60	specified	INTERVALS	13.82
spaces	PANTRIES	20.91	specified	ELSEWHERE	13.64
spaces	DECK	20.18	specified	THEREIN	13.48
spaces	TANKS	19.63	specified	ANNUAL_SURVEYS	13.32
spaces	UNOCCUPIED	19.40	specified	CIRCUMVENTING	13.31
spaces	WATER-BASED	19.27	specified	DEDUCTIBLES	13.31
spaces	WING_TANKS	18.46	specified	ALLOYED	13.31
spaces	POTABLE_WATER	17.84	specified	PASCALS	13.31
spaces	COFFERDAMS	17.55	specified	COMPLETED	12.78
spaces	DOUBLE_HULL	17.53	specified	FUNCTIONS	12.48
spaces	FIRE	17.30	specified	DIMENSIONS	12.42
spaces	PUBLIC	16.97	specified	FEATURES	12.37
spaces	AARE	16.86	specified	OTHERWISE	12.07
spaces	NOISE-ISOLATING	16.86	specified	EXPIRY_DATE	11.57
spaces	OIL-SETTLING	16.86	specified	COLUMN	10.47
spaces	DOUBLE_BOTTOM	16.81	specified	SURVEY	10.21
spaces	PERMEABILITIES	16.74	specified	CONSEQUENT	10.18
spaces	SEMI-ENCLOSED	16.63	specified	THOSE	10.10
spaces	SANITARY	16.12	specified	ADDRESSES	9.79
spaces	PROMENADES	15.74	specified	EXTENT	9.74
spaces	NORMALLY	15.72	specified	INITIAL	9.63
spaces	HOLD	15.62	specified	COLUMNS	9.60
spaces	OCCUPIED	15.58	specified	EQUILIBRIUM	9.60
spaces	SUCH	15.58	specified	DEPOSITARY	9.50
spaces	SIMILAR	15.08	specified	DATES	9.32
spaces	CENTRALIZED	15.07	specified	STEEL_TUBES	8.93
spaces	ENTERING	14.98	specified	SURVIVING	8.93

Source	Target	Score	Source	Target	Score
substances	SELF-REACTIVE	81.89	tank	PORTABLE	137.53
substances	INFECTIOUS	76.52	tank	WAGON	57.61
substances	CLASS	71.40	tank	ROAD	55.44
substances	OZONE	44.69	tank	RAILWAY	53.27
substances	OXIDIZING	44.08	tank	INSTRUCTION	49.18
substances	DEPLETING	43.74	tank	CARGO	34.46
substances	ARTICLES	41.87	tank	WASHING	31.83
substances	PEROXIDES	41.62	tank	WASHINGS	30.55
substances	UNSTABLE	37.41	tank	VENTING	27.68
substances	SPONTANEOUS_COMBUSTION	37.17	tank	MULTIMODAL	27.06
substances	NOXIOUS	33.47	tank	RESIDUE	24.56
substances	DESENSITIZED	32.83	tank	FUEL	23.14
substances	PREPARATIONS	31.84	tank	VEHICLE	22.80
substances	ORGANIC	30.72	tank	SLUDGE	22.33
substances	EVI	28.79	tank	ATTACHED	20.24
substances	CONTAINING	27.76	tank	WASHED	19.42
substances	TOXIC	26.99	tank	ATTEST	19.06
substances	DANGEROUS	26.77	tank	DRAININGS	18.15
substances	EXPLOSIVE	26.45	tank	BELOW-DECK	18.15
substances	HAZARDOUS	24.95	tank	SIGHTING	18.15
substances	TRANSPORTED	24.79	tank	STEAMING	18.15
substances	EXPLOSIVES	24.43	tank	TIE-DOWN	17.71
substances	FORMATION	22.65	tank	GAS-FREEING	17.71
substances	SELF-HEATING	22.14	tank	OIL	17.06
substances	MIXTURES	21.84	tank	FRAMEWORKS	17.01
substances	OZONE-DEPLETING	21.31	tank	SLOPS	16.96
substances	MISCELLANEOUS	20.94	tank	SUPPORTS	16.94
substances	CORROSIVE	19.93	tank	A	16.67
substances	CLASSIFICATION	19.82	tank	GAUGING	16.28
substances	LITERALLY	19.16	tank	VEHICLES	16.12
substances	ELEVATED-TEMPERATURE	19.06	tank	SUMP	16.09
substances	CERTAIN	19.05	tank	MEGC	15.95
substances	INSENSITIVE	18.85	tank	OVERTURNING	15.79
substances	CLASSES	18.65	tank	OILY_WATER	15.62
substances	LIABLE	18.53	tank	DOME	15.62
substances	GROUP	18.09	tank	LIE	15.56
substances	REACTIVE	17.95	tank	CRADLES	15.43
substances	ORGANOMETALLIC	17.93	tank	RETURNED	15.43
substances	ARE	17.92	tank	ENTIRELY	15.41
substances	PYROPHORIC	17.85	tank	PIPING	15.04
substances	WEAPONS	17.65	tank	SEMI-TRAILER	14.89
substances	SOLUTIONS	17.55	tank	EMPTY	14.13
substances	PACKING	17.32	tank	LIFTING	13.57
substances	SPONTANEOUS	17.01	tank	FILLING	13.52
substances	SOLID	16.54	tank	VACUUM-INSULATED	13.51
substances	CONTAIN	16.52	tank	INSTRUCTIONS	13.50
substances	TRANSPORT	16.31	tank	DATA_PLATE	13.48
substances	CLASSIFIED	16.19	tank	ANTI-FLOTATION	13.48
substances	FLAMMABLE_SOLIDS	15.65	tank	HEAVIEST	13.48
substances	DEPLETE	15.33	tank	SERVES	13.36

Source	Target	Score	Source	Target	Score
ballast_water	MANAGEMENT	123.96	bulk_cargoes	SOLID	113.30
ballast_water	SEDIMENTS	117.13	bulk_cargoes	NON-COHESIVE	71.31
ballast_water	EXCHANGE	84.93	bulk_cargoes	IMPLICATIONS	53.25
ballast_water	HAOP	44.05	bulk_cargoes	IMSBC	50.15
ballast_water	SEDIMENT	33.42	bulk_cargoes	GRAIN	33.74
ballast_water	SULLIVAN	32.17	bulk_cargoes	FOREWORD	32.53
ballast_water	HITCH-HIKERS	31.94	bulk_cargoes	DENSITY	31.11
ballast_water	BWE	31.94	bulk_cargoes	ANGLE_OF_REPOSE	29.18
ballast_water	UPTAKE	30.57	bulk_cargoes	CARRIAGE	27.33
ballast_water	EMILY	29.57	bulk_cargoes	COHESIVE	26.02
ballast_water	FEDERAL	27.00	bulk_cargoes	BC	24.40
ballast_water	BMW	27.00	bulk_cargoes	LOADING_PORTS	24.40
ballast_water	SUPRA	26.99	bulk_cargoes	SCHEDULES	24.24
ballast_water	INTRODUCTIONS	26.67	bulk_cargoes	OFFER	22.29
ballast_water	THROUGH	25.81	bulk_cargoes	HIGH	22.14
ballast_water	SHIPS	25.69	bulk_cargoes	SHIPMENT	22.10
ballast_water	GLOBALLAST	25.26	bulk_cargoes	CATEGORIZED	21.81
ballast_water	CONDUCT	24.70	bulk_cargoes	UNLOAD	19.89
ballast_water	MONOGRAPHS	24.61	bulk_cargoes	SIGNIFICANCE	19.89
ballast_water	UPLOADED	24.23	bulk_cargoes	ASPHYXIATION	19.89
ballast_water	STRICTER	24.23	bulk_cargoes	DELEGATES	19.89
ballast_water	SUPPLEMENTAL	24.23	bulk_cargoes	CONSEQUENCE	18.64
ballast_water	NUISANCE	24.07	bulk_cargoes	CONTEMPLATED	18.40
ballast_water	INVADERS	23.98	bulk_cargoes	TRANSPORT_DOCUMENTS	18.40
ballast_water	PUTTING	23.30	bulk_cargoes	POSSESSING	18.29
ballast_water	BOOK	22.77	bulk_cargoes	OXIDATION	17.20
ballast_water	PATHOGENS	21.44	bulk_cargoes	DAMP	17.20
ballast_water	COMBATING	21.32	bulk_cargoes	SINGLE-SIDE	17.20
ballast_water	NOTE	21.06	bulk_cargoes	APPENDIX	17.11
ballast_water	PROMISING	20.91	bulk_cargoes	POTENTIAL	16.39
ballast_water	INTRA-COASTAL	20.91	bulk_cargoes	PRACTICE	16.23
ballast_water	CORY	20.91	bulk_cargoes	SUSCEPTIBLE	16.20
ballast_water	SIMKANIN	20.91	bulk_cargoes	TRANSPORT	15.69
ballast_water	HEBERT	20.91	bulk_cargoes	TWEEN-DECKS	15.36
ballast_water	TREATMENT	20.00	bulk_cargoes	CHARACTERISTIC	15.36
ballast_water	AQUATIC	19.69	bulk_cargoes	HAVING	14.69
ballast_water	DISCHARGE	19.67	bulk_cargoes	SOME	14.10
ballast_water	EPIDEMIC	19.65	bulk_cargoes	INDICATIVE	14.00
ballast_water	DISCHARGES	19.44	bulk_cargoes	ACCEPTABILITY	14.00
ballast_water	SPECIES	19.31	bulk_cargoes	TRIMMING	13.93
ballast_water	ED	18.65	bulk_cargoes	HIGH-DENSITY	12.94
ballast_water	RECORD	18.49	bulk_cargoes	HANDLE	12.08
ballast_water	EXOTIC	18.05	bulk_cargoes	MARITIME	11.36
ballast_water	BUCK	17.07	bulk_cargoes	SECURITY	11.01
ballast_water	MFD	17.07	bulk_cargoes	CODE	10.64
ballast_water	RIDDING	17.07	bulk_cargoes	TRANSPORTING	10.50
ballast_water	LLOYD	17.07	bulk_cargoes	FIRE_PROTECTION	10.25
ballast_water	SLC	17.07	bulk_cargoes	STOWAGE	9.78
ballast_water	OPEN_WATER	17.07	bulk_cargoes	INDIVIDUAL	9.62
ballast_water	GUIDELINES	17.05	bulk_cargoes	SUBSECTION	9.54

Source	Target	Score	Source	Target	Score
bulkhead_deck	BELOW	48.81	dangerous_good	DANGEROUS_GOOD	311.65
bulkhead_deck	LH	38.34	dangerous_good	PHLEGMATIZER	31.14
bulkhead_deck	ABOVE	32.23	dangerous_good	SUSPECTED	25.41
bulkhead_deck	PIERCED	31.26	dangerous_good	IDENTIFIES	23.19
bulkhead_deck	FUELLING	28.93	dangerous_good	SUB-COMMITTEE	22.57
bulkhead_deck	SHELL_PLATING	28.11	dangerous_good	TRANSPORTED	16.64
bulkhead_deck	CRANK	27.09	dangerous_good	POLLUTANT	16.37
bulkhead_deck	SITUATED	23.00	dangerous_good	NATIONS	10.15
bulkhead_deck	PASSENGER	22.23	dangerous_good	SPECIFICALLY	10.07
bulkhead_deck	PARALLEL	22.21	dangerous_good	ONCE	9.91
bulkhead_deck	INOPERABLE	22.10	dangerous_good	NAME	9.54
bulkhead_deck	DRAWN	17.74	dangerous_good	SOURCES	7.50
bulkhead_deck	ESCAPE	17.28	dangerous_good	NORMALLY	7.30
bulkhead_deck	WATERTIGHT_BULKHEADS	14.92	dangerous_good	HAVING	7.20
bulkhead_deck	SPACES	14.74	dangerous_good	MET	6.02
bulkhead_deck	STEPPED	14.43	dangerous_good	LISTED	5.61
bulkhead_deck	FREEBOARD	13.54	dangerous_good	MAY	5.53
bulkhead_deck	GLAND	13.49	dangerous_good	COLUMN	5.37
bulkhead_deck	HAZARDOUS_AREAS	12.71	dangerous_good	PROTECTED	5.29
bulkhead_deck	LINE	12.67	dangerous_good	ESTABLISHED	5.28
bulkhead_deck	EDGE	11.82	dangerous_good	ASSIGNED	5.17
bulkhead_deck	ROUND	11.48	dangerous_good	UNITED	4.57
bulkhead_deck	INACCESSIBLE	11.48	dangerous_good	LIST	4.50
bulkhead_deck	INTERNALLY	11.48	dangerous_good	SUCH	4.17
bulkhead_deck	DRAINED	11.48	dangerous_good	WHERE	4.07
bulkhead_deck	UPPERMOST	11.48	dangerous_good	ENTRY	4.00
bulkhead_deck	HATCH	10.98	dangerous_good	RESPECT	3.99
bulkhead_deck	ACCESSES	10.16	dangerous_good	A	3.95
bulkhead_deck	COLLISION_BULKHEAD	9.49	dangerous_good	GIVEN	3.86
bulkhead_deck	RAMP	9.20	dangerous_good	NUMBER	3.52
bulkhead_deck	VERTICAL_ZONES	8.93	dangerous_good	IT	3.49
bulkhead_deck	THE	8.45	dangerous_good	MARINE	3.48
bulkhead_deck	MOULDED_DEPTH	8.25	dangerous_good	ALSO	3.12
bulkhead_deck	SUBMERGED	8.06	dangerous_good	BE	3.10
bulkhead_deck	WEATHERTIGHT	7.70			
bulkhead_deck	SIDE	7.67			
bulkhead_deck	NON-RETURN	7.39			
bulkhead_deck	EFFECTIVELY	7.37			
bulkhead_deck	TWO	7.35			
bulkhead_deck	IMMERSED	7.11			
bulkhead_deck	NECESSARILY	7.11			
bulkhead_deck	LEAD	6.94			
bulkhead_deck	SHIPS	6.92			
bulkhead_deck	ACCESSIBLE	6.91			
bulkhead_deck	WATERTIGHT_DOORS	6.63			
bulkhead_deck	PLUS	6.63			
bulkhead_deck	HAVING	6.58			
bulkhead_deck	RESTRICT	6.53			
bulkhead_deck	STERN	6.43			
bulkhead_deck	BEING	6.42			

Source	Target	Score	Source	Target	Score
expiry_date	EXISTING	101.81	fire_alarm	FIRE_DETECTION	240.76
expiry_date	UNCHANGED	94.78	fire_alarm	FIXED	77.92
expiry_date	CERTIFICATE	55.79	fire_alarm	SPRINKLER	61.87
expiry_date	BEFORE	46.80	fire_alarm	CENTRALIZED	50.80
expiry_date	BEYOND	43.74	fire_alarm	AUTOMATIC	39.75
expiry_date	MONTHS	39.23	fire_alarm	SYSTEM	39.03
expiry_date	ANNIVERSARY	38.48	fire_alarm	PAGING	36.96
expiry_date	THREE	31.29	fire_alarm	INFORMATICS	36.96
expiry_date	COMPLETED	28.73	fire_alarm	LOBBIES	27.89
expiry_date	REMAIN	23.24	fire_alarm	SYSTEMS	27.54
expiry_date	THE	21.66	fire_alarm	CORRIDORS	24.29
expiry_date	SURVEY	15.47	fire_alarm	BALCONIES	22.06
expiry_date	MAXIMUM	15.43	fire_alarm	EXTRACTION	20.11
expiry_date	PERIOD	14.80	fire_alarm	INSTALLED	17.20
expiry_date	ON_BOARD	14.00	fire_alarm	SPACES	15.03
expiry_date	NEW	13.85	fire_alarm	WHERE	13.91
expiry_date	WITHIN	12.49	fire_alarm	SUBSTITUTE	13.91
expiry_date	SPECIFIED	11.59	fire_alarm	AND	13.04
expiry_date	AFTER	10.04	fire_alarm	SOUNDED	13.00
expiry_date	ALTERING	9.38	fire_alarm	COMPLYING	12.56
expiry_date	DATE	8.58	fire_alarm	INTERMITTENT	12.25
expiry_date	ONE	8.34	fire_alarm	REQUIRED	11.69
expiry_date	VERIFICATION	7.59	fire_alarm	AUDIBLE	11.38
expiry_date	THAN	6.93	fire_alarm	FAN	10.58
expiry_date	MONTH	6.55	fire_alarm	FSS	10.58
expiry_date	LAST	6.34	fire_alarm	CABIN	10.06
expiry_date	CHANGED	6.09	fire_alarm	SIGNALLING	9.78
expiry_date	DAY	5.79	fire_alarm	PANELS	9.14
expiry_date	PROVIDED	5.70	fire_alarm	WATER-SPRAYING	8.86
expiry_date	OF	5.51	fire_alarm	MANUALLY	8.68
expiry_date	SHIP	4.94	fire_alarm	TWO-WAY	8.60
expiry_date	DOCUMENT	4.41	fire_alarm	SIGNALS	8.08
expiry_date	SINCE	4.40	fire_alarm	RADIOTELEPHONE	7.95
expiry_date	FIVE	4.09	fire_alarm	PERIODS	7.82
expiry_date	YEAR	3.69	fire_alarm	SAMPLE	7.79
expiry_date	MAY	3.49	fire_alarm	VISUAL	7.50
expiry_date	EXCEEDING	3.47	fire_alarm	CONNECTED	7.05
			fire_alarm	CONTAINING	6.61
			fire_alarm	INITIATED	6.61
			fire_alarm	FUNCTION	6.40
			fire_alarm	INDICATOR	6.39
			fire_alarm	IDENTIFIABLE	6.19
			fire_alarm	INSTALLATION	6.10
			fire_alarm	DEPENDING	6.09
			fire_alarm	ELSEWHERE	6.00
			fire_alarm	SO	5.95
			fire_alarm	SPACE	5.90
			fire_alarm	DETECTION	5.55
			fire_alarm	DOOR	5.30
			fire_alarm	CAPABLE	5.20

Source	Target	Score	Source	Target	Score
freight_containers	OVERPACKS	119.94	general_purpose_container	OPEN_TOP_CONTAINER	146.92
freight_containers	ERIES	70.24	general_purpose_container	VARIETY	73.44
freight_containers	ISO	62.64	general_purpose_container	ROOF	46.42
freight_containers	LASHINGS	45.94	general_purpose_container	RESPECTS	32.39
freight_containers	TRILINGUAL	40.55	general_purpose_container	GREATEST	29.93
freight_containers	MIXED_LOADS	40.55	general_purpose_container	SIMILAR	24.41
freight_containers	LIMITB	40.55	general_purpose_container	CONTAINER	20.43
freight_containers	GENERAL_PURPOSE_CONTAINERS	40.55	general_purpose_container	OPENING	17.71
freight_containers	TELEX	40.55	general_purpose_container	ALLOW	16.46
freight_containers	SUFFICE	40.55	general_purpose_container	VENTILATED	13.69
freight_containers	BULKAND	40.55	general_purpose_container	CAPABILITY	11.05
freight_containers	PACKAGES	37.61	general_purpose_container	AIR	10.21
freight_containers	CONVEYANCES	36.24	general_purpose_container	POSSIBLE	9.34
freight_containers	TANK_CONTAINERS	33.07	general_purpose_container	DESIGNED	8.45
freight_containers	TR	30.61	general_purpose_container	CONTAINERS	8.43
freight_containers	RATIONALE	28.66	general_purpose_container	EXCEPT	7.74
freight_containers	SERIES	28.03	general_purpose_container	HAVING	7.62
freight_containers	SMALL	27.05	general_purpose_container	A	7.50
freight_containers	CONTAINERS	26.47	general_purpose_container	GENERAL	7.47
freight_containers	SPECIFICATION	24.74	general_purpose_container	PURPOSE	7.46
freight_containers	CEDEX	23.38	general_purpose_container	BUT	7.19
freight_containers	ACCUMULATION	22.47	general_purpose_container	IS	6.21
freight_containers	LARGE	21.70	general_purpose_container	TYPES	6.03
freight_containers	CONTAINER_SHIPS	20.24	general_purpose_container	CARRYING	5.57
freight_containers	SERVICEABILITY	20.24	general_purpose_container	CAPABLE	4.22
freight_containers	CODING	18.09	general_purpose_container	BULK	3.86
freight_containers	PURPOSE-BUILT	18.09	general_purpose_container	ALL	3.75
freight_containers	ROAD_VEHICLES	18.09			
freight_containers	SECURING	16.86			
freight_containers	GENERAL_CARGO	16.80			
freight_containers	RO-RO_SHIPS	16.51			
freight_containers	INTERMODAL	15.27			
freight_containers	TERMINOLOGY	15.27			
freight_containers	CLOSED	14.71			
freight_containers	CSI	14.28			
freight_containers	SIFTPROOF	14.28			
freight_containers	INSERT	14.28			
freight_containers	TI	13.45			
freight_containers	CORNER	13.38			
freight_containers	FISSILE	13.14			
freight_containers	FREIGHT_CONTAINER	13.01			
freight_containers	POINTS	12.98			
freight_containers	EDITION	12.84			
freight_containers	IN_TRANSIT	12.75			
freight_containers	CHASSIS	12.75			
freight_containers	INDEXES	11.17			
freight_containers	RATINGS	10.56			
freight_containers	EXCLUSIVE_USE	10.38			
freight_containers	LSA-I	9.74			
freight_containers	TABLE	9.69			

Source	Target	Score	Source	Target	Score
gross_tonnage	UPWARDS	128.67	international_convention_on_load_lines	ADDITIONS	50.26
gross_tonnage	WORLD'S	56.01	international_convention_on_load_lines	SIDESCUTTLE	43.58
gross_tonnage	FIFTY	53.18	international_convention_on_load_lines	MODIFIED	39.22
gross_tonnage	SHIPS	47.65	international_convention_on_load_lines	FORCE	31.60
gross_tonnage	ABOVE	38.97	international_convention_on_load_lines	FORWARD_PERPENDICULAR	25.09
gross_tonnage	LESS	37.49	international_convention_on_load_lines	DEFINED	20.27
gross_tonnage	PERCENT	35.12	international_convention_on_load_lines	FREEBOARD	15.91
gross_tonnage	THAN	32.60	international_convention_on_load_lines	DEADLIGHTS	15.35
gross_tonnage	CENT	31.82	international_convention_on_load_lines	RELATING	14.51
gross_tonnage	MERCHANT	30.43	international_convention_on_load_lines	MEANINGS	14.46
gross_tonnage	NEW_SHIPS	29.33	international_convention_on_load_lines	AS	12.52
gross_tonnage	EVERY	29.24	international_convention_on_load_lines	GANGWAY	12.51
gross_tonnage	MERCHANT_SHIPPING	26.32	international_convention_on_load_lines	NEW_SHIPS	12.51
gross_tonnage	THIRTY-FIVE	24.95	international_convention_on_load_lines	LENGTH	12.16
gross_tonnage	LOOKS	24.95	international_convention_on_load_lines	DRAUGHT	11.95
gross_tonnage	TONS	24.70	international_convention_on_load_lines	WHATEVER	11.57
gross_tonnage	NEAR-COASTAL	24.63	international_convention_on_load_lines	ANNEXES	11.38
gross_tonnage	MATES	24.12	international_convention_on_load_lines	BREADTH	11.28
gross_tonnage	ENGAGED	21.24	international_convention_on_load_lines	HEREINAFTER	10.72
gross_tonnage	TWENTY-FIVE	20.24	international_convention_on_load_lines	SILL	10.48
gross_tonnage	CALCULATED	20.00	international_convention_on_load_lines	LONDON	9.14
gross_tonnage	CONSTRUCTED	19.79	international_convention_on_load_lines	ADOPTED	8.76
gross_tonnage	DISTINCTIVE	19.05	international_convention_on_load_lines	NON-RETURN	8.43
gross_tonnage	GOVERNMENT'S	17.64	international_convention_on_load_lines	THE	8.34
gross_tonnage	BUT	16.85	international_convention_on_load_lines	DONE	8.32
gross_tonnage	PASSENGER	16.74	international_convention_on_load_lines	INLETS	8.27
gross_tonnage	VOYAGES	16.69	international_convention_on_load_lines	AGREED	8.15
gross_tonnage	WORLD	16.58	international_convention_on_load_lines	FREEBOARD_DECK	8.12
gross_tonnage	OIL_TANKER	16.13	international_convention_on_load_lines	PROTOCOL	8.02
gross_tonnage	GT	15.70	international_convention_on_load_lines	LIMITING	7.97
gross_tonnage	OF	15.67	international_convention_on_load_lines	IN	7.27
gross_tonnage	TONNAGE	15.55	international_convention_on_load_lines	ZONES	7.23
gross_tonnage	PORT_OF_REGISTRY	15.13	international_convention_on_load_lines	EXEMPTED	6.84
gross_tonnage	OVER	14.35	international_convention_on_load_lines	PROVISIONS	6.69
gross_tonnage	MATE	13.49	international_convention_on_load_lines	EQUILIBRIUM	6.58
gross_tonnage	AND	12.81	international_convention_on_load_lines	SIDESCUTTLES	6.58
gross_tonnage	PER	12.48	international_convention_on_load_lines	DEFINITIONS	6.31
gross_tonnage	TANKERS	11.88	international_convention_on_load_lines	TEXT	6.01
gross_tonnage	MORE	11.87	international_convention_on_load_lines	REGULATION	5.83
gross_tonnage	EXISTING_SHIPS	11.62	international_convention_on_load_lines	DISCHARGES	5.82
gross_tonnage	FLEET	11.62	international_convention_on_load_lines	STRENGTH	5.72
gross_tonnage	SHIP	11.52	international_convention_on_load_lines	FOLLOWS	5.46
gross_tonnage	OIL_TANKERS	11.33	international_convention_on_load_lines	BULK_CARRIERS	4.70
gross_tonnage	SEAGOING	11.08	international_convention_on_load_lines	BULK_CARGO	4.61
gross_tonnage	NAME_OF_SHIP	11.02	international_convention_on_load_lines	PERMITTED	4.59
gross_tonnage	EXEMPTED	10.98	international_convention_on_load_lines	PARTIES	4.51
gross_tonnage	CARGO	10.04	international_convention_on_load_lines	EXISTING	4.49
gross_tonnage	SHARE	9.61	international_convention_on_load_lines	DETERMINED	4.02
gross_tonnage	WHEREVER	9.22	international_convention_on_load_lines	ARTICLES	3.91
gross_tonnage	CRAFT	9.11	international_convention_on_load_lines	HOWEVER	3.80

Source	Target	Score	Source	Target	Score
liquefied_gases	NON-REFRIGERATED	110.01	machinery_spaces	CATEGORY	75.27
liquefied_gases	REFRIGERATED	96.56	machinery_spaces	CONTROL_STATIONS	49.49
liquefied_gases	CERTIFICATE_OF_FIT NESS	94.67	machinery_spaces	AUTO-TELEPHONE	40.27
liquefied_gases	IGC	40.44	machinery_spaces	GALLEYS	36.33
liquefied_gases	TEMPRATURES	35.04	machinery_spaces	PUMP-ROOMS	35.46
liquefied_gases	NONREFRIGERATED	35.04	machinery_spaces	TURRET	33.47
liquefied_gases	NON-OXIDIZING	35.04	machinery_spaces	TREADS	32.88
liquefied_gases	NON-FLAMMABLE	34.95	machinery_spaces	CROWNS	32.88
liquefied_gases	BULK	25.60	machinery_spaces	LAUNDRIES	32.82
liquefied_gases	GASESFOR	24.75	machinery_spaces	ACCOMMODATION	27.90
liquefied_gases	REFRIGERATING-MACHINE	24.75	machinery_spaces	CONTIGUOUS	26.30
liquefied_gases	LOW-PRESSURE	24.71	machinery_spaces	UNATTENDED	25.85
liquefied_gases	CHEMICALS	22.61	machinery_spaces	IMPENDING	23.25
liquefied_gases	BOILING	22.29	machinery_spaces	UNSUPERVISED	23.25
liquefied_gases	CARRIAGE	19.58	machinery_spaces	FIXED-PRESSURE	23.25
liquefied_gases	TRANSPORT	18.35	machinery_spaces	SPACES	23.20
liquefied_gases	CLASS	15.37	machinery_spaces	INTERNAL_COMBUSTION	21.78
liquefied_gases	COMPRESSED	13.72	machinery_spaces	DRYING	20.64
liquefied_gases	HIGH_PRESSURE	13.18	machinery_spaces	FIRE-EXTINGUISHING	20.12
liquefied_gases	FLAMMABLE	11.85	machinery_spaces	OIL-FIRED	19.23
liquefied_gases	FOR	11.61	machinery_spaces	ROOMS	19.18
liquefied_gases	LAYING	11.60	machinery_spaces	WATER-BASED	17.88
liquefied_gases	CGA	11.60	machinery_spaces	FACE	16.50
liquefied_gases	TRANSPORTED	11.34	machinery_spaces	ACHINERY	16.41
liquefied_gases	DUE	10.66	machinery_spaces	STEAM_TURBINES	16.41
liquefied_gases	INTRODUCE	10.48	machinery_spaces	FLOOR_PLATING	16.41
liquefied_gases	UNSTABLE	10.48	machinery_spaces	RECESSED	16.41
liquefied_gases	REACTIONS	10.02	machinery_spaces	STOREROOMS	16.41
liquefied_gases	CHEMICALLY	10.02	machinery_spaces	LOCKABLE	16.41
liquefied_gases	INCLUDES	9.81	machinery_spaces	PERIODICALLY	15.55
liquefied_gases	NON-TOXIC	9.27	machinery_spaces	WORKSHOPS	15.05
liquefied_gases	SURGE	9.27	machinery_spaces	BOUNDARIES	14.86
liquefied_gases	LEGIBLY	9.27	machinery_spaces	TRUNKS	13.76
liquefied_gases	PORTABLE	9.00	machinery_spaces	SERVICE	13.51
liquefied_gases	PRESSURE	8.55	machinery_spaces	ENGINE_ROOMS	13.37
liquefied_gases	REGARD	8.37	machinery_spaces	RELAXING	13.37
liquefied_gases	ATMOSPHERIC_PRESSURE	7.92	machinery_spaces	SEPARATES	13.37
liquefied_gases	CERTAIN	7.88	machinery_spaces	PURIFIER	13.37
liquefied_gases	MIXTURES	7.80	machinery_spaces	SEA-GOING	13.30
liquefied_gases	ROAD	7.34	machinery_spaces	VENTILATION	13.15
liquefied_gases	CRITICAL	7.02	machinery_spaces	SYSTEMS	12.92
liquefied_gases	DISSOLVED	6.87	machinery_spaces	BILGE_PUMPS	12.77
liquefied_gases	THERMALLY	6.73	machinery_spaces	MANNED	12.49
liquefied_gases	FILLING	6.30	machinery_spaces	CARGO	12.34
liquefied_gases	POINT	6.29	machinery_spaces	INCINERATORS	12.29
liquefied_gases	HYDROGEN	6.15	machinery_spaces	ESCAPE	11.61
liquefied_gases	ACCESSORIES	6.14	machinery_spaces	ELECTRIC_MOTORS	11.56
liquefied_gases	PROTOTYPE	6.04	machinery_spaces	ILLUMINATED	11.56
liquefied_gases	MANIFOLD	6.04	machinery_spaces	CASINGS	11.48
liquefied_gases	TOXIC	5.55	machinery_spaces	SERVING	11.33

Source	Target	Score	Source	Target	Score
medical_first_aid_guide	ACCIDENTS	149.00	noxious_liquid_substances	VARIOUSLY	62.94
medical_first_aid_guide	INVOLVING	119.93	noxious_liquid_substances	CATEGORIZATION	44.48
medical_first_aid_guide	USE	42.20	noxious_liquid_substances	BULK	32.86
medical_first_aid_guide	CARGO-SPECIFIC	41.26	noxious_liquid_substances	HIGH-VISCOSITY	25.67
medical_first_aid_guide	MEDICAL_GUIDE	35.72	noxious_liquid_substances	CARRY	24.61
medical_first_aid_guide	FIRST_AID	32.62	noxious_liquid_substances	CARRIAGE	18.75
medical_first_aid_guide	FIRST-AID	21.51	noxious_liquid_substances	CERTIFIED	18.29
medical_first_aid_guide	MFAG	17.28	noxious_liquid_substances	RESIDUES	16.08
medical_first_aid_guide	EMS	17.28	noxious_liquid_substances	PROBABLE	15.68
medical_first_aid_guide	REFER_TO	16.26	noxious_liquid_substances	NLS	14.77
medical_first_aid_guide	RECENT	12.56	noxious_liquid_substances	CALLING	14.77
medical_first_aid_guide	ILO	12.56	noxious_liquid_substances	LIQUEFIED_GAS	14.50
medical_first_aid_guide	MEDICAL	12.40	noxious_liquid_substances	MIXTURES	13.54
medical_first_aid_guide	EDITION	11.36	noxious_liquid_substances	IDENTIFIED	12.90
medical_first_aid_guide	FOR	11.04	noxious_liquid_substances	DISCHARGES	12.07
medical_first_aid_guide	TRANSPORT_DOCUMENT	10.94	noxious_liquid_substances	LISTING	11.82
medical_first_aid_guide	IMO	9.34	noxious_liquid_substances	CHEMICALS	11.42
medical_first_aid_guide	BOARD	8.98	noxious_liquid_substances	CARRIERS	11.13
medical_first_aid_guide	IN	7.66	noxious_liquid_substances	VISCOSITY	11.04
medical_first_aid_guide	CARRYING	7.66	noxious_liquid_substances	POLLUTION	10.95
medical_first_aid_guide	DETAILED	6.29	noxious_liquid_substances	WASHINGS	10.71
medical_first_aid_guide	MOST	5.40	noxious_liquid_substances	DISCHARGED	9.80
medical_first_aid_guide	AMENDED	5.10	noxious_liquid_substances	INCIDENT	9.60
medical_first_aid_guide	RECOMMENDATIONS	4.90	noxious_liquid_substances	CERTIFICATE	9.38
medical_first_aid_guide	DANGEROUS_GOODS	4.47	noxious_liquid_substances	DISCHARGE	8.48
medical_first_aid_guide	CONTAINED	4.21	noxious_liquid_substances	CONTAINING	8.21
medical_first_aid_guide	ORGANIZATION	4.15	noxious_liquid_substances	PETROLEUM	8.15
medical_first_aid_guide	WHO	3.46	noxious_liquid_substances	MARINE	7.87
medical_first_aid_guide	REFERENCE	3.18	noxious_liquid_substances	RESULTING	7.35
medical_first_aid_guide	COMMITTEE	3.00	noxious_liquid_substances	HOSES	6.99
			noxious_liquid_substances	IN	6.96
			noxious_liquid_substances	UNLOADING	6.60
			noxious_liquid_substances	PLAN	6.56
			noxious_liquid_substances	CATEGORY	6.49
			noxious_liquid_substances	DIVIDED	5.89
			noxious_liquid_substances	AMOUNTS	5.78
			noxious_liquid_substances	OIL	5.42
			noxious_liquid_substances	DEFINED	5.27
			noxious_liquid_substances	CATEGORIES	5.17
			noxious_liquid_substances	EMERGENCY	5.13
			noxious_liquid_substances	DRAWN	5.02
			noxious_liquid_substances	INTO	4.88
			noxious_liquid_substances	LISTED	4.84
			noxious_liquid_substances	ISSUED	4.72
			noxious_liquid_substances	FOUR	4.71
			noxious_liquid_substances	TERMINAL	4.56
			noxious_liquid_substances	TERMINALS	4.51
			noxious_liquid_substances	COVER	4.40
			noxious_liquid_substances	DANGEROUS	4.07
			noxious_liquid_substances	EXCESS	4.01

Source	Target	Score	Source	Target	Score
oil_record_book	PART	47.88	oil_tankers	TONNES	73.56
oil_record_book	RECORDED	30.21	oil_tankers	DEADWEIGHT	73.00
oil_record_book	INSPECT	22.17	oil_tankers	DELIVERED	65.40
oil_record_book	OIL_FILTERING	14.66	oil_tankers	RUDE	64.86
oil_record_book	MADE	13.92	oil_tankers	SBT	56.52
oil_record_book	ENTRY	12.27	oil_tankers	SATISFYING	52.42
oil_record_book	ENTRIES	11.67	oil_tankers	RODUCT	35.48
oil_record_book	DELAY	10.95	oil_tankers	BUILT	28.12
oil_record_book	REFERRED	9.16	oil_tankers	REGS	25.59
oil_record_book	KEPT	8.48	oil_tankers	CBT	22.10
oil_record_book	OPERATIONS	8.27	oil_tankers	CARRIER	21.12
oil_record_book	BALLAST	7.26	oil_tankers	BULK_CARRIERS	20.17
oil_record_book	CIRCUMSTANCES	7.14	oil_tankers	JUNE	19.41
oil_record_book	HOLDING	7.06	oil_tankers	DESIGNATED	19.13
oil_record_book	COMPLETED	6.91	oil_tankers	PRODUCT	19.04
oil_record_book	NOTED	6.70	oil_tankers	COW	18.14
oil_record_book	SHIP'S	6.51	oil_tankers	CARRIERS	17.67
oil_record_book	IN	6.30	oil_tankers	APPLY_TO	17.38
oil_record_book	SHALL	6.16	oil_tankers	REGULATION	17.22
oil_record_book	BE	5.71	oil_tankers	CUSTODY	16.65
oil_record_book	AN	5.43	oil_tankers	IMPORTING	16.65
oil_record_book	TAKING	5.36	oil_tankers	DE-BALLAST	16.65
oil_record_book	ANNEX	4.93	oil_tankers	ASPHALT	16.65
oil_record_book	INSPECTION	4.19	oil_tankers	NON-OIL-CARGO	16.65
oil_record_book	ON_BOARD	4.18	oil_tankers	CONVERSIONS	16.65
oil_record_book	WITHOUT	4.12	oil_tankers	SINGLE-HULL	16.65
oil_record_book	THE	3.97	oil_tankers	OIL	15.71
oil_record_book	EVENT	3.58	oil_tankers	DOUBLE_HULL	14.11
oil_record_book	WHETHER	3.46	oil_tankers	ANGULAR	13.57
oil_record_book	SO	3.07	oil_tankers	OCTOBER	13.50
			oil_tankers	BEFORE	13.20
			oil_tankers	METERS	13.17
			oil_tankers	SEGREGATED_BALLAST	12.93
			oil_tankers	DEDICATED	12.43
			oil_tankers	WASHING	12.20
			oil_tankers	INSTALLED	11.98
			oil_tankers	DWT	11.73
			oil_tankers	EMULSION	11.73
			oil_tankers	INTERPRETATIONS	11.57
			oil_tankers	GROSS_TONNAGE	11.33
			oil_tankers	LINERS	10.96
			oil_tankers	MONITORING	10.69
			oil_tankers	TANKERS	10.67
			oil_tankers	CRUISE	10.66
			oil_tankers	DENY	10.47
			oil_tankers	GOAL-BASED	10.47
			oil_tankers	CLEAN_BALLAST	9.88
			oil_tankers	UNIFIED	9.86
			oil_tankers	ABOVE	9.81
			oil_tankers	KEELS	9.77

Source	Target	Score	Source	Target	Score
on_board	SHIP	43.98	rescue_boats	FAST	208.47
on_board	ASHORE	35.26	rescue_boats	CERTIFICATE_OF_PROFICIENCY_IN_SURVIVAL_CRAFT	77.27
on_board	SEEMP	30.48	rescue_boats	SELF-BAILING	44.59
on_board	PERSONS	25.58	rescue_boats	PROFICIENCY	38.19
on_board	SHIPS	24.50	rescue_boats	INFLATED	37.78
on_board	RETENTION	23.51	rescue_boats	SURVIVAL	35.73
on_board	CARE	22.40	rescue_boats	SURVIVAL_CRAFT	32.46
on_board	RETAINED	21.88	rescue_boats	STEERED	27.30
on_board	SEAFARERS	18.84	rescue_boats	MOBILITY	27.30
on_board	COMMITTED	17.42	rescue_boats	BOAT-HOOK	27.30
on_board	DOCKSIDE	17.14	rescue_boats	BAILING	27.30
on_board	KEPT	16.93	rescue_boats	CRAFT	26.12
on_board	CARGO_UNITS	16.00	rescue_boats	LIFEBOATS	23.30
on_board	INSPECT	15.71	rescue_boats	FENDERS	22.27
on_board	MEDICAL	15.15	rescue_boats	CANDIDATE	21.74
on_board	PRE-SEA	14.74	rescue_boats	HOLDER	17.17
on_board	ACCOUNTABLE	14.74	rescue_boats	CLEARED	15.72
on_board	OILY_MIXTURES	14.46	rescue_boats	ATTEND	14.54
on_board	TIMING	14.44	rescue_boats	OPERATIVE	13.59
on_board	POSTED	14.44	rescue_boats	SELF-RIGHTING	13.59
on_board	MANIFEST	14.03	rescue_boats	ALLOWING	12.92
on_board	EXPIRY_DATE	13.99	rescue_boats	CAPSIZE	12.81
on_board	EXISTING_SHIP	13.81	rescue_boats	INFLATABLE	12.73
on_board	FIRE_EXTINGUISHERS	12.82	rescue_boats	BOAT	12.55
on_board	PLACED	12.39	rescue_boats	HEADWAY	12.14
on_board	KEEP	12.32	rescue_boats	SURVIVORS	12.14
on_board	CARRIED	12.28	rescue_boats	EVERY	11.69
on_board	CARRY	12.25	rescue_boats	ENGINES	9.96
on_board	PRESENTING	12.18	rescue_boats	DRAINAGE	9.46
on_board	END-RESULT	12.12	rescue_boats	THAN	8.92
on_board	FIRE-DETECTION	12.12	rescue_boats	DEMONSTRATE	8.58
on_board	LF	12.12	rescue_boats	COMPETENCE	7.81
on_board	RESOLVES	12.12	rescue_boats	WEATHERTIGHT	7.76
on_board	SPENDS	12.12	rescue_boats	OTHER	7.73
on_board	SECURING_CARGOES	12.12	rescue_boats	SPECIFICATION	7.68
on_board	MARINE_TERMINALS	12.12	rescue_boats	PASSENGER	7.61
on_board	NON-METHANE	12.12	rescue_boats	ADVANCED	7.31
on_board	EMERGENCYPLAN	12.12	rescue_boats	RUN	7.04
on_board	RELATIVES	12.12	rescue_boats	LOG-BOOK	7.04
on_board	INTENSIFYING	12.12	rescue_boats	REQUIREMENTS	7.02
on_board	BAY_PLAN	12.12	rescue_boats	STOP	6.92
on_board	RECONSTITUTION	12.12	rescue_boats	LAUNCHING	6.81
on_board	SMALLPOX	12.12	rescue_boats	ACCESSORIES	6.80
on_board	UNCERTAINTY	12.12	rescue_boats	FRESH_WATER	6.80
on_board	CEASING	12.12	rescue_boats	SHALL	6.71
on_board	ILLNESSES	12.12	rescue_boats	BOATS	6.69
on_board	OOK	12.12	rescue_boats	AUTOMATICALLY	6.66
on_board	ILL	12.12	rescue_boats	STOWAGE	6.65
on_board	RESIDUAL_FUEL_OIL	12.12	rescue_boats	BUOYANT	6.48
on_board	ADOPTED	12.12	rescue_boats	REQUIREMENT	6.33

Source	Target	Score	Source	Target	Score
steering_gear	ELECTROHYDRAULIC	74.42	survival_craft	EMBARKATION	52.28
steering_gear	FIREMEN'S	46.42	survival_craft	LAUNCHING	42.63
steering_gear	ELECTRIC	41.15	survival_craft	UNSERVICEABLE	41.37
steering_gear	POWER	40.06	survival_craft	TRANSPONDER	39.26
steering_gear	RECOGNIZABLE	37.88	survival_craft	BOATS	36.09
steering_gear	POWER_FAILURE	37.88	survival_craft	RESCUE	35.31
steering_gear	MAIN	36.53	survival_craft	RESCUE_BOATS	32.46
steering_gear	CIRCUIT	35.54	survival_craft	STATIONS	29.89
steering_gear	COMPARTMENT	34.60	survival_craft	GLOWING	29.28
steering_gear	WINDLASS	32.82	survival_craft	MOTORIZED	29.28
steering_gear	LINKAGE	32.82	survival_craft	SART	29.28
steering_gear	CONTROLLERS	32.82	survival_craft	DROGUE	29.28
steering_gear	BUSBARS	32.79	survival_craft	EQUIPPING	29.28
steering_gear	DEMANDS	29.32	survival_craft	REFLECTOR	29.25
steering_gear	RUDDER	25.78	survival_craft	SURVIVORS	27.71
steering_gear	AUXILIARY_STEERING_GEAR	21.81	survival_craft	RESCUE_BOAT	26.97
steering_gear	SUPPLYING	20.87	survival_craft	LOST	21.85
steering_gear	REMOTE	19.97	survival_craft	MOUNT	20.68
steering_gear	XI	19.71	survival_craft	ABANDONMENT	20.59
steering_gear	CONTROL	19.37	survival_craft	CLOSEST	19.44
steering_gear	NAVIGATING	19.31	survival_craft	RADAR	17.91
steering_gear	TRANSMITTERS	18.91	survival_craft	FAST	16.87
steering_gear	ANGULAR	18.91	survival_craft	INTERPRET	16.87
steering_gear	COMPRISE	18.14	survival_craft	DAVIT	16.87
steering_gear	RUDDER_STOCK	16.85	survival_craft	ANTENNA	16.14
steering_gear	WINCHES	16.37	survival_craft	MANAGE	15.54
steering_gear	STABILIZER	16.37	survival_craft	LAUNCH	15.01
steering_gear	PROVEN	16.30	survival_craft	AIS	15.01
steering_gear	STEERING	15.34	survival_craft	STOWED	14.75
steering_gear	BRIDGE	15.24	survival_craft	CLEARING	14.59
steering_gear	DISCONNECTING	14.62	survival_craft	TRANSPONDERS	14.59
steering_gear	STEER	14.62	survival_craft	TRANSMITTER	14.59
steering_gear	TORQUE	14.62	survival_craft	LIFEJACKET_LIGHTS	14.59
steering_gear	RESERVOIR	14.62	survival_craft	ABANDONING	14.59
steering_gear	OUTFITS	14.56	survival_craft	BOAT	14.25
steering_gear	ACTUATORS	13.34	survival_craft	STOWAGE	13.74
steering_gear	FED	13.34	survival_craft	WHEREBY	13.68
steering_gear	AUXILIARY	13.32	survival_craft	SURVIVAL	13.34
steering_gear	POWER-OPERATED	12.49	survival_craft	APPLIANCE	13.34
steering_gear	IMPROVED	12.34	survival_craft	DAVIT-LAUNCHED	13.30
steering_gear	STOCK	12.34	survival_craft	CRAFT	12.13
steering_gear	APPURTENANCES	11.53	survival_craft	FLOAT-FREE	11.89
steering_gear	WATERTIGHT_DOORS	11.44	survival_craft	DON	10.99
steering_gear	HYDRAULIC	11.28	survival_craft	ATTEND	10.99
steering_gear	ALARMS	11.28	survival_craft	SEA-ANCHOR	10.99
steering_gear	SWITCHBOARD	10.68	survival_craft	MUSTER	10.90
steering_gear	COMPRISING	10.33	survival_craft	RENDERED	10.52
steering_gear	CHANGE-OVER	10.29	survival_craft	DONNED	10.27
steering_gear	GEAR	10.07	survival_craft	BRINGING	10.27
steering_gear	ROOMS	9.98	survival_craft	FAILURE	10.26