جــامــعــة الـــشـرق الأوسـط
MIDDLE EAST UNIVERSITY

Evaluating Semantic Measures by Computing the Coverage of a Condensed Text

تقييم المقاييس الدلالية من خلال قياس تغطية النص المُكثف

BY

Student Name: Zainab Mahmoud Bayram

Student Number: 401320077

Supervisor

Prof. Ahmad K. A. Kayed

This thesis submitted in partial fulfillment of the Requirements for Master Degree in Computer Science

Department of Computer Science

Faculty of Information Technology

Middle East University

January, 2016

AUTHORIZATION STATEMENT

I, Zainab Mahmoud Bayram, authorize the Middle East University to provide hard copies or soft copies of my thesis to libraries, institutions or individuals upon their request.

Name: Zainab Mahmoud Bayram**.**

Data: 4/1/2016

Signature: 

إقرار تفويض

أنا زينب محمود بيرم ، أفوض جامعة الشرق الأوسط للدراسات العليا بتزويد نسخ من رسالتي ورقياً وإلكترونياً للمكتبات أو المنظمات أو الهيئات و المؤسسات المعنية بالأبحاث و الدراسات العليا عند طلبها.

الأسم: زينب محمود بيرم

التأريخ: 2016/1/4

التوقيع: 

Middle East University

Examination Committee Decision

This is to certify that the thesis entailed "Evaluating Semantic Measures by Computing the Coverage of a Condensed Text" was successfully defended and approved on, 2016.

Examination Committee Members	**Signature**
Prof. Ahmad Kayed **(supervisor & Member)** Dean of Faculty of Information Technology Middle East University (MEU)	
Dr. Sadeq AlHamouz **(Internal Member)** Associate professor, Faculty of information technology Middle East University (MEU)	
Prof. Bilal Anas Abu AlHuda **(External member)** Department of Computer Science Yarmouk University	B. Abu Huda

Acknowledgment

First of all, I would like to express my deepest appreciation to my supervisor Prof. Ahmad Kayed for his guidance and advice through my research.

Special thanks to my parents, sisters and brothers for their continuous support.

Finally, I would like to express my gratitude to my dear friend Hadeel Alkazaz for her support, encouragement and patience.

بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

"وقل ربي زدني علماً"

Dedication

I would like to exploit this opportunity to dedicate this project to my father, mother, brothers and sisters, without whose invaluable support. I would not have been able to have achieved this in my lifetime.

May God bless them.

Table of Contents

LIST OF TABELS

LIST OF FIGURES

LIST OF ABBRIVATIONS

CC	:	Condensed Concepts.
DUC	:	Document Understanding Conference.
HSO	:	Hirst & St-Onge.
IC	:	Information Content.
ICF	:	Information Content Family.
KAON	:	Karlsuhe Ontology.
LCH	:	Leacock & chodorow.
LCS	:	Least Common Subsumer.
MSE	:	Mean Square Error.
NIST	:	National Institute of Standards and Technology.
NLP	:	Natural Language Processing.
OC	:	Original Concepts.
PLF	:	Path Length Family.
SM	:	Semantic Measures.
SRF	:	Semantic Relatedness Family
WS4J	:	WordNet Similarity for Java.
WuP	:	Wu and Palmer.

Evaluating Semantic Measures by Computing the Coverage of a Condensed Text

Prepared by: Zainab Mahmoud Bayram

Supervisor: Prof. Ahmad K. A. Kayed

Abstract

Automatic summarization systems that produce a condensed text are an essential topic in natural language processing (NLP). How to evaluate these systems is an issue for many researchers. There are several approaches to define the quality of a condensed text. This thesis used semantic measures to define the quality of a condensed text. Semantic measures compute the similarity and relatedness among concepts included in knowledge source. This thesis used semantic measures to find how much the condensed text cover the original text.

This thesis applied several experiments to find which semantic measure or measures best cover a condensed text. Well-known benched data sets have been collected with both original text and condensed text for those experiments. The main concepts for both the original and the condensed text have been extracted using ontological tools. Six semantic measures from three families (i.e. path, information contents and relatedness) have been applied to those extracted concepts; those measures are (WuP, LCH, Resnik, LIN, HSO and LESK). Above 10, 000 data items for 48 files for two data set groups have been used to find out and evaluate which semantic measure or measures are best to compute how much a condensed text covers its original. This thesis used mean square error to compute the difference between the evaluation of the proposed semantic measures and the expert evaluation. The results showed that from six similarity measures, Resnik was the best measure that identified the quality of a condensed text where it gave minimum MSE (15.6%) from the expert evaluation with different acceptance rates. This thesis found that LCH has the minimum MSE (12%) in most cases if we assume that all condensed text have 100% coverage for the original text.

In the case of bad summaries i.e. where the expert stated that a condensed text has less than 40% coverage of the original text; in that case, this research recommended using LESK measure as it gave minimum MSE (3%). This thesis showed that the semantic measures can be used to identify not only the good coverage but also the bad coverage.

Keywords: Ontology, Concept, condensed text, extracted concepts, semantic similarity, semantic relatedness.

تقييم المقاييس الدلالية من خلال قياس تغطية النص المُكثف

إعداد: زينب محمود بيرم

أشراف: أ. د. أحمد الكايد

المُلخص

نظم التلخيص التلقائي التي تنتج النص الملخص من اهم الموضوعات البحثية في معالجة اللغة و تقييم نوعية النصوص الملخصة. هناك عدة طرق لتحديد نوعية النصوص الملخصة، منها على سبيل المثال: أهمية الجملة بالنسبة للنص، عدد مرات استدعاء الجملة في النص الملخص. هذه الرسالة تستخدم المقاييس الدلالية لتحديد نوعية النص الملخص، وقياس مدى تشابه و تقارب المفاهيم المعرفية. تهدف هذه الرسالة الى معرفة مقدار تغطية النص الملخص للنص الاصلي باستخدام المقاييس الدلالية.

من خلال بحثنا، تم اجراء العديد من التجارب لايجاد اي مقياس/ مقاييس الدلالية المغطية للنص الاصلي بأفضل نسبة. تم جمع مجموعة من البيانات التي تحتوي على النصوص الاصلية و النصوص الملخصة. تم أستخراج المفاهيم الاساسية من كلا النصين الأصلي و الملخص. تم تطبيق ستة مقاييس دلالية من ثلاثه أنواع (المسار، محتوى المعلومات، القرابة) على هذه المفاهيم المستخرجة، حيث تم تطبيقها على 48 ملف من البيانات لمعرفة اي المقايسس الدلالية التي تعطي أفضل النتائج من حيث تغطية النص الملخص.

هذه الرسالة تستخدم متوسط مربع الخطأ لحساب الفرق بين تقييم المقاييس الدلالية وتقييم الخبراء. أظهرت النتائج ان من خلال استخدام ستة مقاييس دلالية، ان ريسنك (Resnik) كان افضل مقياس الذي يحدد نوعية النص الملخص بنسبة أقل خطأ (15.6%). كما اوجدت الدراسة ان مقياس ال سي اج (LCH) كان افضل مقياس لتحديد نوعية النص الملخص اذا افترضنا ان النص الملخص مغطى النص الاصلي بنسبة 100% وكانت نسبة الخطأ (12%). في حالة كان الملخص سيْ اي عند تقيم الخبير لتغطية النص الملخص وجد ان النص الملخص يغطي 40% من النص الاصلي. من خلال التجارب وجدنا ان مقياس ليسك(LESK) اعطانا نسبة خطأ (3%). من خلال البحث تبين ان المقاييس الدلالية يمكن استخدامها ايضا في معرفة الملخص السيء.

كلمات البحث: علم الوجود، المفاهيم، النص الملخص، المفاهيم المستخرجة، التشابه الدلالي،الارتباط الدلالي.

CHAPTER ONE

Introduction

1.1 Introduction

The importance of a text summarization system has been growing with the expansion of the quantity of online information. Condensing knowledge is the aim of many organizations. The data (original text) may be summarized in more than one concise form. The goal of automatic summarization is to take information source for a particular domain, and extract content from it while preserving the main information content of the original text and represent contents to the user in an abstract, condensed structure forming a text summary.

Automatic summarization research becomes an important topic in natural language processing (NLP), at the same time we need to discuss and clarify the issues on how to evaluate the text summarization systems. According to Steinberger and Jezek, the evaluation of summary quality is very hard and challenging task. There are several measures used to define the quality of a condensed text. Examples of these measures are sentence recall, sentence ranking and question answering. Different types of measures have been used to find which measure is the best measure (Steinberger and Jezek, 2009; House, et al., 2002).

1.1.1 Automatic Summarization and Evaluation

The main goal of automatic summarization is to take original text, extract the main content from it then presents the most important information to the user in condensed form. In general, summaries can be user focus (topic focus or query focus) which determine the contents as the users require or it could be generic which locate the main content covered by the original data.(Mani,2001a ;Mani, 2001b ; Alguliev & Aliguliyev, 2007).

Maintaining coherence in summary is important too, created a summary based on cutting & pasting the text can cause a problem on coherence among sentences. There are many ways to classify text summaries based on different criteria; Sparck-Jones classified summaries as below: (Sparck-Jones, 1999 as cited on Fukusima & Okumura 2001*)*.

- Input factors: text length, genre, single vs. multiple documents.
- Purpose factors: who is the user, the purpose of summarization?
- Output factors: running text or headed text etc.

Evaluation text summarization is crucial. There are several challenges in evaluating summaries such as summaries are a result from (NLP), in some cases it is difficult to know what is the correct output. This is the same as machine translation output. Most machines generated summaries need an expert evaluation to judge the summary. Therefore, this may increase the expensive of evaluation. Normally, we can judge the relevance of the summaries by applying different methods which can be classified into two categories: intrinsic and extrinsic evaluating measures. (1) Intrinsic evaluation methods are based on analysis the summary or assess the coherence and informativeness of the summaries. (2) Extrinsic evaluation tests the summaries based on another task like reading comprehension, question answering. This research used intrinsic evaluation to define the coverage of a condensed text (Gupta, 2014).

Kayed et al. introduced a coverage measure to define the quality of description for specific domain knowledge where the higher coverage value indicates a better quality for description (Kayed, 2013).

1.1.2 Ontology and Semantic Measures

Jiang et al. defined the Ontology as "an abstract description system for knowledge composition in a certain domain". Also, they added that ontology supplies a standardized vocabulary for representing entities in the domain. Ontologies can be classified in their purpose as: general purpose ontologies and domain specific ontologies. "Many researches are using WordNet[1] as ontology" (Jiang et al., 2013).

WordNet

WordNet is an online lexical which is created by the Cognitive Science Laboratory at Princeton University, it can be seen as an ontology. WordNet contains nouns, verbs, adjectives, and adverbs. It contains sets of synonymous word senses which are known as synsets. Version WordNet 2.0 contains 115,000 different synsets. These synsets contain 80,000 nouns, 13,500 are verbs, 18,500 are adjectives and 3700 are adverbs. Each synset may contain one or more synonymous word. Also, each synset has brief definition "gloss" to define the meaning of the synsets. WordNet also defines the relationships between each synset. A relation between each synset is known as the semantic relation, and the relation between the word senses is known as the lexical relation. The semantic relation is the relation between each two synsets such as (hypernym, hyponym, meronym, and holonym) relations (Michelizzi, 2005; Boon young & Mingkhwan 2015).

[1] https://wordnet.princeton.edu/

The relation hypernym/hyponym is known as is-a relationship. For example, an orange is a fruit. The relation meronym/holonym is known as part-of relationship. For example, a mouse is part of computer. The most common relation in WordNet is hypernym/hyponym (is-a) which is considered as 80% of the relations. In general, the hypernym/hyponym relation is considered about how two concepts are similar. The meronym/holonym relation is considered as how two concepts are related to each other (Meng et al., 2013).

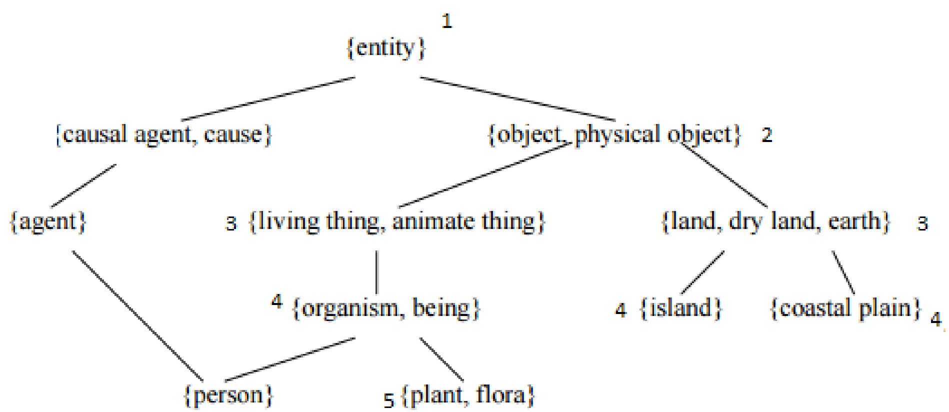


Figure 1.1: WordNet Hypernyms adapted from (Michelizzi, 2005)

To explain the relation clearly, Figure 1.1 shows a sample for WordNet hypernyms between nouns. For example, (land, dry land, earth) is the hypernym of (island) because an island is a land. Hyponym relation is the opposite of hypernym relation; this means that island is a hyponym of (land, dry land, earth). Figure 1.1 structure shows that the deeper concepts are the more specific and vise versa the upper concepts are more abstract. Thus, (plant, flora) is more specific than (living thing, animate thing). Therefore, the most abstract concept is entity which is considered as the root of the taxonomy (Meng et al., 2013).

Semantic Measures

Semantic measures can be classified into two groups: measures of semantic similarity and measures of semantic relatedness. Figure 1.2 shows the classification of semantic measures (Michelizzi, 2005). Each group contains a number of measures. We will use different measures from each type to find how much the condensed text covered the original text. Slimani, state that “semantic similarity and semantic relatedness are two related, but semantic similarity is more specific than relatedness and can be considered as a type of semantic relatedness. For example ’student’ and ‘professor’ are related terms but not similar (Pederson, et al. 2004; Slimani, 2013).

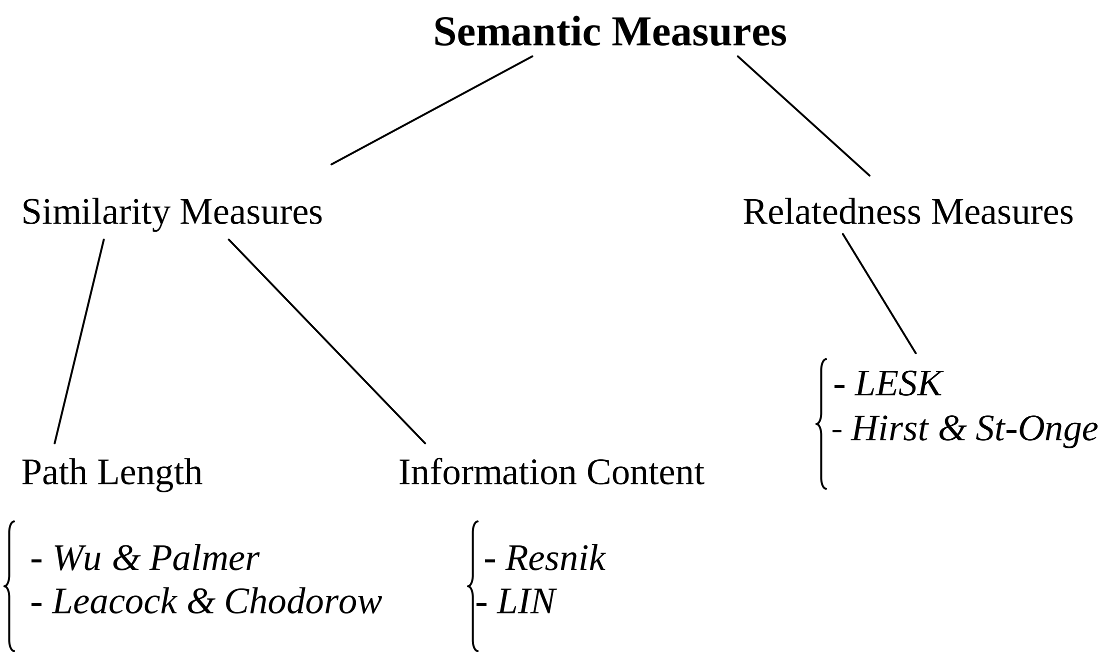


Figure 1.2: Taxonomy of Semantic Measures adapted from (Michelizzi, 2005)

Many ontological tools can be used to extract concepts from the text. For example: KAON[2], Swoogle[3], and Portégé[4]. KAON is an ontology management infrastructure targeted for business applications. It includes a comprehensive tool suite allowing easy ontology creation, storage, retrieval and maintenance of ontologies.” Swoogle computes the rank of each semantic web documents and provides an online system to check the availability of ontologies in any domain. While Portégé is another tool that allows a user to construct domain ontology, customize data entry form, and enter data. This research uses KAON due to the availability, ease of use, and it has an efficient user interface”. This research used KAON to extract the main concepts from both the original text and the condensed (Kayed, et al. 2013).

1.2 Problem Statement

Automatic summarization is crucial and it is important for many domains. Knowledge summary (condensed text) is the aim of many knowledge extractors. There are several measures to define the quality of a condensed text such as sentence recall, sentence ranking and question answering. This research deploys the semantic measures to find the coverage between condensed text and original text. Semantic measures (similarity measures and relatedness measures) compute the similarity and relatedness between concepts included in knowledge source. This research assumes that the good condensed text is the text that has "good" coverage for the original text. This research used semantic measures to find how much the condensed text cover the original text. It also defines which measure or measures best compute the coverage for a text.

---

[2] http://kaon2.semanticweb.org/
[3] http://swoogle.umbc.edu/
[4] http://protege.stanford.edu/

1.3 Problem Statement Questions

- Which semantic measures can measure the coverage of a condensed text?
- How these semantic measures can be used to define the coverage for a condensed text?
- Which semantic measure is the best to evaluate a condensed text?

1.4 Limitations

This research collected data sets and used different semantic measures. This research limited its collected data to the following two datasets: The first dataset is a benchmark from NIST, which contains 34 original texts with its condensed text. The original text sizes vary from 4KB to 20KB with a condensed text size 1KB. The second dataset contains 14 original texts with its condensed text. The original text sizes vary from 3KB to 19KB while condensed text come with size 1KB. This research used KAON to extract the concepts from the original text and condensed text with default frequency (3). This research used six semantic measures (WuP, LCH, Resnik, LIN, HSO, and LESK).

1.5 Objectives

Evaluation the condensed text is an essential issue in summarization. This research aims to evaluate the condensed text. There are several measures have been used to compute the similarity or relatedness of concepts. This research deploys the semantic measures to evaluate the quality of condensed text. We used existing semantic measures to define the quality of the condensed knowledge. The main aim is to give an efficient evaluation of all these measures and finds the best measure that can give the minimum error that define the coverage of the condensed text.

1.6 Motivation

There are many available tools for text summarization. Finding the best tool that can give the coverage among original text and condensed text is a challenge. In summarization as a fast development area, there is a need for finding a proper evaluation methodology. Many of research address summary evaluation by applying different evaluation methods to measure the quality of a condensed text. There is a lack of measurers that compute the quality of the condensed text. We need to find which measure from the semantic measures that best compute the coverage for an original text.

1.7 Contribution

Evaluating the text summarization is very important issue. This research finds how to evaluate the quality of the condensed text using semantic measures. This research contributes to defining the best semantic measures with minimal error that defines how much a condensed text covers its original text.

1.8 Organization of the Thesis

The thesis contains five chapters, references, and appendices. The following part explains a brief description for each chapter:

Chapter 2 presents a theoretical background and literature about the ontology, the importance of summarization and how important to evaluate the summaries. This chapter also contains an introduction about semantic measures.

Chapter 3 presents the proposed model steps. Its explain how to extract the concepts from condensed text and original text. Also describe the different types of semantic measure and how the concepts matching process are done using these measures.

Chapter 4 explains the experiments in details. Its present the matching process in details and how we define how much the condensed text cover its original text using semantic measures.

Chapter 5 present thesis discussion, conclusion and future work.

CHAPTER TWO

Literature Review & Related Works

Overview

This chapter presents a theoretical background and literature that relates to our study, we classified the literature into five parts: the first section is about automatic summarization and how important to evaluate the summaries. The second section discusses semantic measures and how it can be used in different domains to compute the relations between concepts and terms. The third section discusses the ontology and WordNet. The fourth part clarifies different types of semantic similarity and relatedness measures. Last section explain in brief summarization evaluation types.

2.1 Automatic Summarization Evaluation

In the following, we provide a brief idea about the automatic text summarization and why summary evaluation is important. It gives a background about methods used in evaluation, as each method has different measures. This research focus on semantic measures that can be classified into two parts: semantic similarity measures and semantic relatedness measures.

Jing, et al. used two methods to evaluate the automatic summarization systems which are: an evaluation of generated summaries against an ideal summary and evaluation of how well summaries help a person performs a task such as information retrieval. They carried two large experiments for both kinds. Their focus was on how different factors can be affected on the final evaluation results, for example, summary length. They found that summary length affects on both types. On the “ideal” summary based evaluation, accuracy decrease as summary length increase. While for the other type of evaluation they found that summary length and accuracy on information retrieval task appeared to correlate randomly (Jing, et al. 1998).

Mani discussed different methods for evaluation automatic summarization systems. Hence, he lists several serious challenges in evaluation summaries such as automatic summarization is a machine producing output, in some cases the output may be correct but in others it's hard to arrive at notation of what the correct output is. He also classified the evaluation text summarization methods in two categories. The first method is the intrinsic evaluation, which is based judging the relevance of the condensed text by matching it with reference summary generated by a human. The second one is the extrinsic evaluation, which depends on completion of some task like reading comprehension. He also discussed some measures that to compare between different summaries such as sentence recall measure and sentence ranking measure (Mani, 2001a).

Gupta discussed how important the evaluation of text summary. The author states that we can judge any summary by the relevance of summary. This is done by applying extrinsic and intrinsic evaluation measures. The intrinsic evaluation techniques judge the summary relevance with human evaluation. Different factor discussed like maintaining coherence in summary, maintaining information in summary, calculating Recall and Precision, Ranking of lines, similarity of contents. Extrinsic evaluation techniques judge the quality of summary by performing some tasks on the summary like: game of questions and game of classification (Gupta, 2014).

2.2 Semantic Measures

This part discusses different types of semantic measures and the importance of semantics measures in many domains. Some of the researchers focuses on similarity measures while other focus on relatedness measures. Several studies use both measures to find which measure that gives the best result.

Pedersen, et al. discussed the semantic similarity and semantic relatedness on WordNet. Using similarity measures can tell how much two concepts or terms are alike. For example, the automobile might considered more like a boat than a tree. Similarity measures are restricted on pairs of nouns, verbs, adjectives, and adverbs. However, measures of relatedness can measures how pairs of concepts are related to each others. For example, murder and gun are related. Because of this, different measures of relatedness can be applied on wider area comparing it with similarity measures (Pedersen, et al. 2004).

Michelizzi classified the semantic measures into two main groups: measures of semantic similarity and measures of semantic relatedness. Semantic similarity measures work on noun-noun or verb-verb pairs using is-a hierarchies while semantic relatedness measures work on all open part of speech because there are not limited to is-a hierarchies. Semantic similarity measures can be classified in different families: the first one is path length similarity measures which are also known as the edge (or node) counting measures. These measures quantify similarity according to the length of the shortest path (example: Wu & Palmer). The second one is information content based measures; these measures compute similarity as a function of the information content (IC) of the most specific common subsumer in the ontology (example: Resnik) (Michelizzi, 2005).

Buckley, et al. Proposed a new semantic similarity measure based on hybrid methods. They discussed the importance of semantic similarity measures in many fields. They focus on WordNet based semantic similarity measures which can be classified into three categories: first category is node-based methods which use information content(IC) to compute the amount of information contained in WordNet. Second category is edge-based methods which calculate the edge length between two concepts to find the shortest path between them. Third category is hybrid methods which use a combination of information from different resources to compute the similarity between concepts. The new semantic similarity used the internet as a corpus and the structure information from WordNet because they believe that the internet is the largest source and it can regularly be updated (Buckley, et al. 2011).

Slimani focuses on semantic similarity measures approaches. The author uses two widely used benchmarks. He classified the similarity measures into different methods. First method is structured - based measures which based on computing the path length between two concepts. Second method is information content measures based on information content to measure how two concepts are similar to each other. Information contents value depends on the frequency of concepts in a text. Third method is feature-based measures assume that each concept is defined by a set of concepts. Fourth method is hybrid measures combine between approaches such as information content based and path based measures. Several measures from each approach being compared to give an efficient evaluation of all the measures. Although the author focus on semantic similarity measures but he also discuss ontologies used on semantic similarity and proposed several examples based on its purpose (general purpose ontologies or domain specific ontologies) (Slimani, 2013).

McInnes & Pedersen evaluated the measures if semantic similarity and relatedness in biomedical text. They measure how two concepts are similar or related to each other based on classification: similarity measures which are classified to path measures and information content measures and relatedness measures. The paper used MSH-WSD dataset which is provided from national library of medicine. Their focus was to evaluate the efficiency of these measures and find that the information content (IC) measures can give a higher accuracy than the other measures (McInnes & Pedersen, 2013).

Mittal & Jain studied the problem in query expansion, where most of the times user's query may contain unclear terms which add relevant and irrelevant terms to the query. They present a method to improve this problem by using semantic similarity and relatedness measures between the ambiguous terms. They apply different type of measures such as Leacock & Chodorow and Wu & Palmer similarity measures on noun only as most of the information is represented by nouns (Mittal & Jain, 2015).

Al-Khiaty & Ahmed reviewed the matching model as its an essential in many model management operations such as model evaluation and retrieval. The authors focus in software development models as for each software system there is a set of models that describes its structural, behavioral and functional perspectives. They identified the matching between models and finding the similarities and differences in each one, especially in UML diagrams (Unified Modeling Language) class diagram. They use semantic similarity to compare concepts according to WordNet in class diagram where concepts here are (classes's names, operation's names, attributes' names, and the relation between classes). Semantic similarity used is based on semantic path-based measures. They used two measures supported by WordNet: path length and Wu & Palmer measures (Al-Khiaty & Ahmed, 2015)

2.3 Ontology

Ontology is an abstract description system for knowledge composition in a certain domain. It can give a description of concepts or terms in an effective way. Ontologies can be used in different domains; each domain has its own vocabularies, concepts or terms. Many researchers classified the ontologies based on their purpose.

Kayed et al. discussed the ontology importance in many domains. Ontology includes vocabularies of concepts and specification of their meaning. It also can improve understanding in how the concepts or terms are related to each others. They state that ontologies are used in many domains such as: artificial intelligence, software engineering, semantic web, biomedical informatics and library science. They focus on building ontology in software engineering domain. They developed new ontology in requirement engineering process using KAON's tool. This will enable developers to share a common concepts and terms and allow them to understand the domains in simple language (Kayed, et al. 2010).

Seddiqui. & Aono defined ontology base on Gruber definition for ontology "ontology is an explicit specification of a conceptualization". They classified the ontologies based on their size, small- scale or large-scale. Often, large-scale ontologies represent distributed knowledge area within a problem domain (Seddiqui, M. H., & Aono, M. 2010).

2.4 Semantic Similarity and Relatedness Measures

According to literature, semantic measures can be classified in difference families based on their theoretical principles. We choose different measures to cover all kinds of semantic measures. The following will list two measures from each family.

1. Path Length Family
   - Wu and Palmer Measure.
   - Leacock & Chodorow Measure.
2. Information Content Family
   - Resnik Measure.
   - LIN Measure.
3. Semantic Relatedness Family
   - LESK Measure.
   - Hirst & St-Onge Measure.

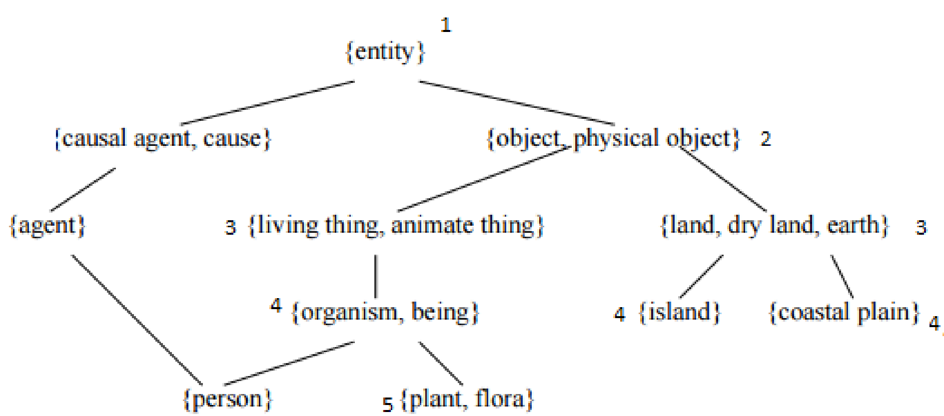


Figure 2.1: WordNet Hypernyms adapted from (Michelizzi, 2005)

2.4.1 Path Length Family

Path length similarity measures (also known as edge or node counting measures) compute the similarity between two concepts as a function of the length of the path linking the concepts and on the position of the concepts in the taxonomy. For example, (plant, flora) is closer to (living thing, animate thing) than it is to (land, dry land, earth). It can be seen as undirected graph. The greater the distance between two concepts, the less similar they are. Well known measures are Wu and Palmer and Leacock & Chodorow. Wu and Palmer similarity metric measure compute the depth of the two concepts while Leacock & Chodorow measure finds the shortest path between two concepts or terms using node counting (Michelizzi, 2005).

Wu and Palmer Measure

Michelizzi illustrate WuP similarity measure which is the depth of the two concepts and the depth of the least common subsumer (LCS).

The shared parent of two synsets is called subsumer. Baader et. al discussed LCS and shows that if we have two concepts, each concept represented by a node and both concepts shared the same ancestor. This relation is defined by is-a relationship. For example we can say that a car is an automobile and automobile is a vehicle. Also we can say boat is a vehicle. LCS and WuP measure are related to each other where the deeper the LCS is the larger the value of the measure (Baader et. al, 2007).

For example figure 3.4 the subsumers are (object, physical object) and (entity) for nodes (living thing, animate thing) and (land, dry land, earth). But to find the least common subsumer for these two nodes; we must search for the most specific subsumer of the two synsets. Therefore, the LCS for these two synsets is (object, physical object) since its more specific than entity (Yang, 2015; Michelizzi, 2005).

$$Sim_{wup} = \frac{2 * depth(LCS)}{depth(concept_1) + depth(concept_2)} \quad \text{.................................. (1)}$$

Equation (1) shows how to calculate the WuP measure, which is the node depth of LCS for the two nodes divided by the sum of the depth of first node and the depth of the second node. Figure 3.4, to found the similarity of the two nodes (island) and (coastal plain) using WuP; the node counting (island) and (coastal plain) is 4 for both, the depth for their LCS which is (land, dry land, earth) is 3. Thus, the score using equation (1) is $\frac{2*3}{4+4}$=0.75.

Leacock & Chodorow Measure

LCH measure is another measure that uses the depth and the distance using nodes counting.

Equation (2) shows how to calculate the LCH measure.

$$Sim_{lch} = -log(\frac{dist_{node}(s_1, s_2)}{2 \cdot D}) \quad \text{... (2)}$$

Where S1 is the first concept, S2 is the second concept; dist is the distance between S1 and S2. D is the depth for a given taxonomy where the concepts are existing.

For example, using figure 3.4, the two synsets (island) and (coastal plain). The distance between them is 3, and the depth is 4. Thus, the score using measure LCH by equation (2) is:

$-\log 3/(2 * 4) \approx 0.9808$

2.4.2 Information Content Family

Information Content (IC) measures use the information content of concepts to measure the semantic similarity between two concepts. The information content value of the concept is calculated based on the frequency of the concepts, the concepts that occurs a lot have low information content. The concepts that have high information content are the concepts that rarely occur (Slimani, 2013).

Mathematically, the information content for a given concept can be calculated as equation (3):

$$IC(c) = -logP(c) \qquad \text{......} (3)$$

Where P(c) is the probability of the concept c. “High information content means that the concept conveys a lot of meaning when it occurs in a text. A concept with high information content is very specific, but a concept with low information content is very general; therefore, information content corresponds to specificity” (Michelizzi, 2005).

Resnik Measure

Resnik measure is information content measure. It took into account the LCS information content which return the information content of the LCS of two concepts.

Equation (4) shows how to calculate Resnik measure.

$$Sim_{res} = IC(LCS) \qquad \text{...} (4)$$

LIN Measure

LIN measure which is building on Resnik's measure of similarity. To calculate the similarity between two concepts using LIN measure; the more these two concepts are similar to each other, the more they will have in common. When the two concepts are exactly the same concept, a LIN measure result is the maximum similarity.

Equation (5) shows how to calculate LIN measure. By equation (5), we can note that the similarity based on the information content for the least common subsumer. And the information content for both concepts. LIN measure and WuP measure look alike, but the WuP measure based on the depth of the LCS, where LIN measure based on the information content of LCS (Corley & Mihalcea, 2005; Michelizzi, 2005).

$$Sim_{lin} = \frac{2 * IC(LCS)}{IC(concept_1) + IC(concept_2)} \qquad \text{..} (5)$$

(Slimani, 2013; Baader et. al, 2007)

2.4.3 Semantic Relatedness Family

Semantic relatedness is a much broader notion than semantic similarity. For example, a tire is related to car, but the two are not very similar since a tire is not a type of a car nor is a car a type of tire. Well known measures are LESK and Hirst & St-Onge measures.

LESK Measure

LESK measure finds the relatedness of two concepts by defining a function of the overlapping between the corresponding definitions provided by a dictionary. It's based on the glosses of the synsets, where the synset that has gloss that contain a common words, they are more related to each other (Michelizzi, 2005).

Hirst & St-Onge Measure

HSO measure classifies relations in WordNet as having directions. "Its establishes the relatedness between two concepts by trying to find a path between them that is neither too long nor that changes directions too often" (Pedersen et al., 2004; Liu et al., 2011).

2.5 Summarization Evaluation Types

As discussed in previous sections, summarization is important in many fields like information retrieval. It can save time by reading the summary instead of reading the whole original text. Also it can speed up the information retrieval. The evaluation of summary quality is important and challenging task. Table 2.1 shows examples of methods used to evaluate the quality of the condensed text. These methods like sentence recall used exact matching among the original text and the condensed text. This research used semantic matching to evaluate the quality of the condensed text. To apply the semantic matching technique; we used semantic measures which can be classified into semantic similarity and semantic relatedness.

Table 2.1: Example of Methods used to Evaluate the Condensed Text

Type of Evaluation	Author	Method	Procedure
Intrinsic	Jing et. al	Sentence recall	Using sentence recall to measure how much the summary contains the sentences from the text.
Intrinsic	Mani	Sentenced ranking	The summary is specified in term of ranking the sentences in terms of worthiness

Extrinsic	Morris et. al	Question answering	Also known as reading comprehension task. In this task human first read the full original text or summary for a given document, then the human answer a test of multiple question tests. By scoring the percentage of the correct answers.

2.6 Tools Used

Many tools have been used in this research. Below is a brief description of each tool.

2.6.1 KAON Text2Onto Tool

KAON Text2Onto is a tool support the ontology engineering process by text mining techniques; we used this tool to extract the main concepts from each original text and condensed text. Figure 2.1 shows the front end of text2Onto. Each file is converted to text file and then uploaded it to get the concepts from both. (Maedche A., 2001)

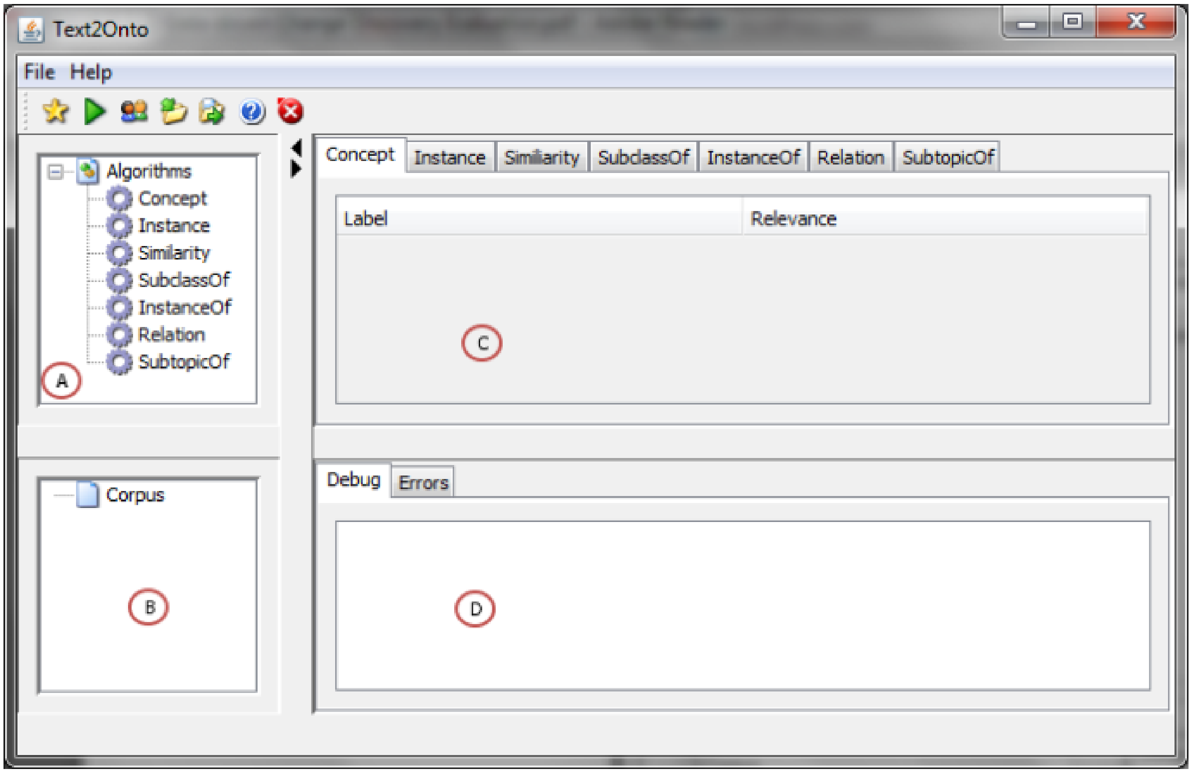


Figure 2.2: KOAN Text2Onto Front End

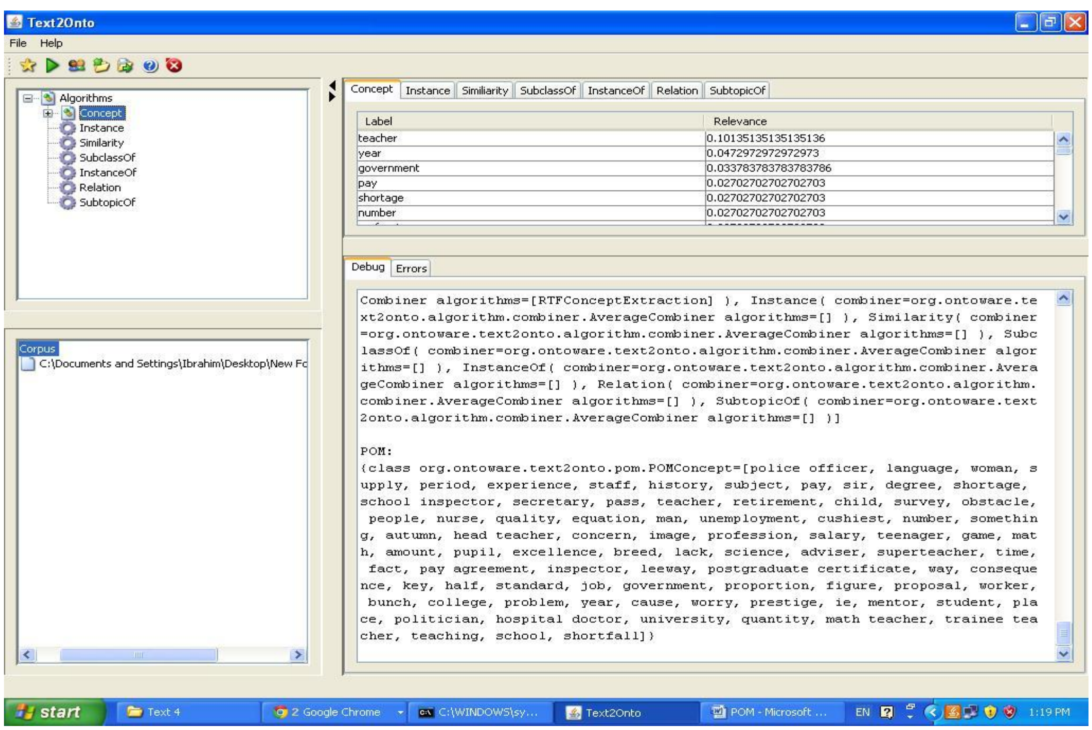


Figure 2.3: Concepts Extraction Example using KOAN Text2Onto

2.6.2 WordNet Similarity for Java (WS4J)

This tool WS4J[5] provides a pure Java API for several semantic similarity and semantic relatedness measurements. The results are depends on WordNet relations between pairs of concepts. Figure 2.3 shows front end of WS4J. It provides matching for different semantic measures such as WuP, LCH, Resnik, LIN, HSO, and LESK.

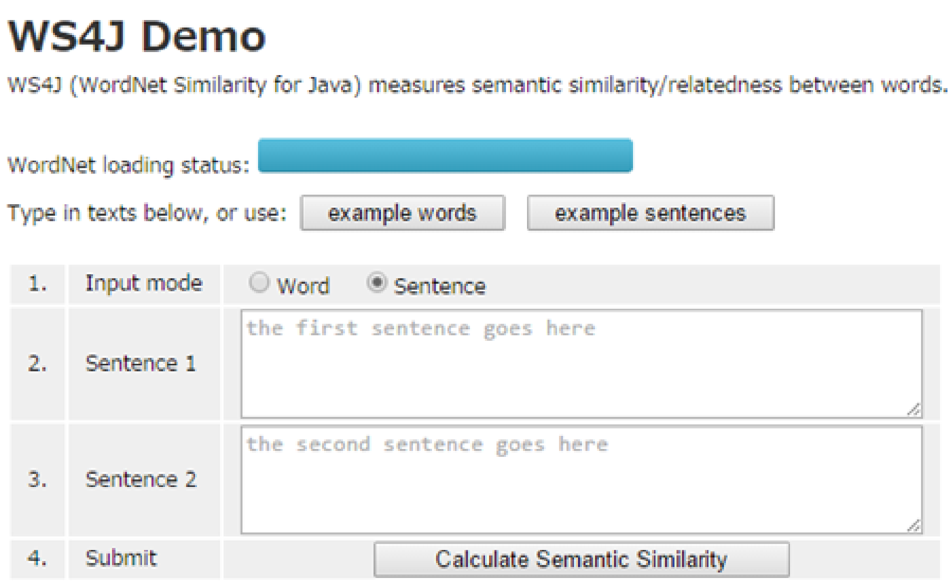


Figure 2.4: WS4J Tool Front End

WS4J provide two options of matching. First, by matching only two pairs of concepts and compute the semantic results of each measure. Second option, by matching set of concepts with another set of concepts and then computes the semantic results. This option can save time and effort; because it calculates the result for a number of concepts at once.

---

[5] http://ws4jdemo.appspot.com/

CHAPTER THREE

Data Collection and Concepts Extraction

Overview

This chapter explains the methodology that the researcher has used. The steps of each phase of the methodology will be detailed. These steps are: data collection, concepts extraction, applying different semantic similarity and relatedness on the extracted concepts and evaluate the results of each measure by computing the error of each one. The aim is to match between original text and condensed text using coverage technique.

3.1 Introduction

The methodology of this research combined the quantitative and qualitative approach. Our methodology has been based on building several experiments to find the best semantic measure. The experiment part of the proposed work will be considered as quantitative. To be able to know the quality of the condensed text, we need human judgments. Thus, part of the evaluation process has been based on human. The other part is done by our experiment and this will depends on the error calculation which is the difference between the human result and measures results.

The main idea of this research is to find which semantic measures that give us the quality of the condensed text by extracting the main concepts for both condensed text and original text using KAON software; then matching those concepts by applying different types of semantic measures, through these measures the coverage will be defined. According to Kayed et al. using different measures will enable us to obtain “good concepts”. These concepts are not too generalized concepts neither too specified one (Kayed et al., 2013).

To evaluate our approach, an expert needed to evaluate the quality of condensed texts and these results will be compared with our results to see how much we are close to expert evaluation.

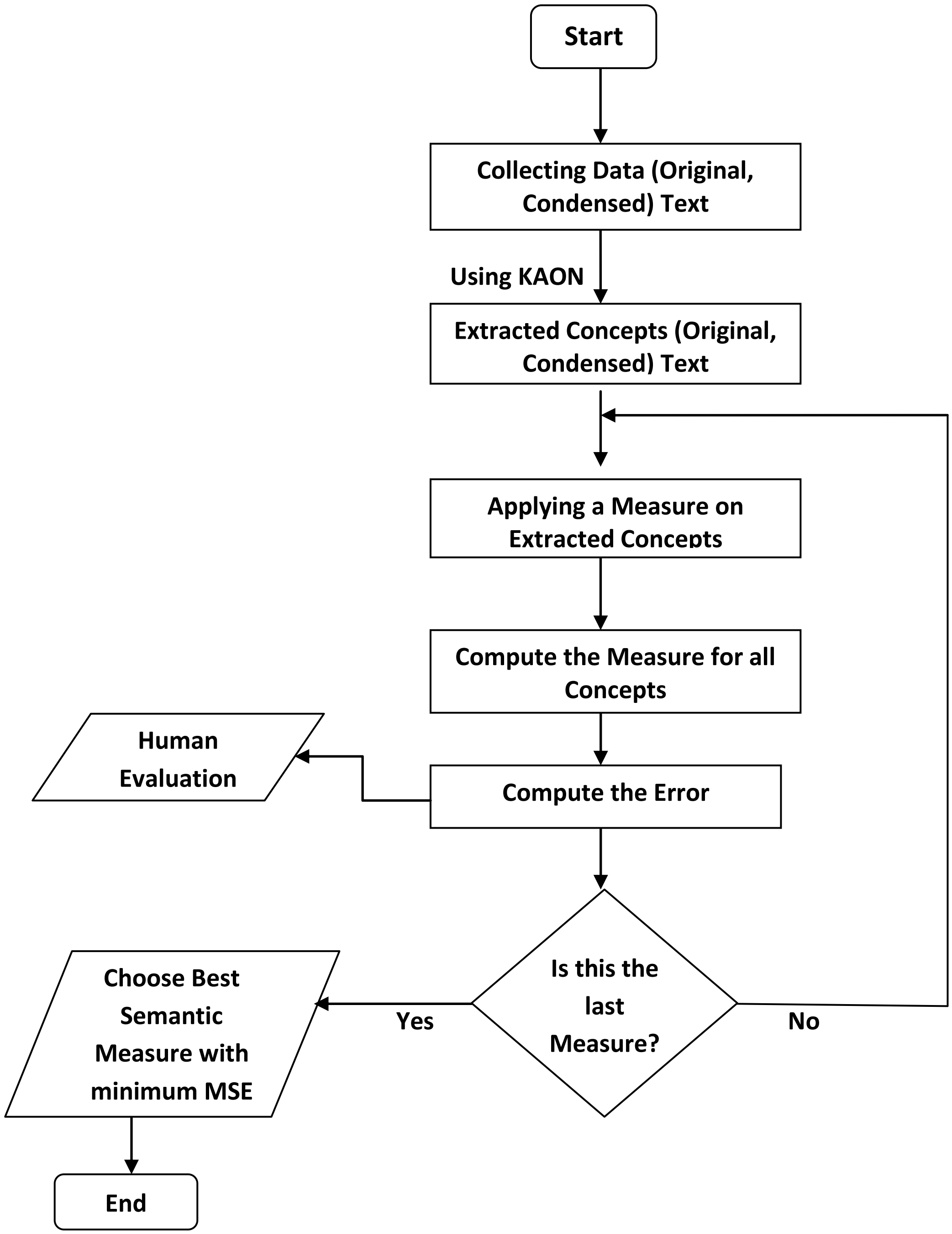


Figure 3.1: Flowchart of the Proposed Solution

The following will illustrate the main steps of the research methodology as showing in figure 3.1:

1- Collecting datasets.

2- Extracting the concepts.

3- Apply several semantic measures.

4- Evaluate the results.

The methodology will contain the following steps in details:

3.2 Data Collection (Data Samples)

To be able to apply our experiment, good data sets are needed. At the beginning we thought to use technical paper and its abstract, but we found that their abstract is very short and may not represent the whole paper. Therefore, we decide to look for a data where a condensed text and original text has been collected.

This research used a well-known benchmark (data samples) released on 2001 by NIST (National Institute of Standards and Technology) which is DUC[6] (Document Understanding Conference). It contains seven datasets (DUC2001 to DUC2007). Lin & Moghaddas et al. discuss the validity of evaluation methods used in DUC. To get this data set, an assigned agreement is required. We sent the agreement to NIST and got the DUC datasets. *Refer to Appendix 1.*

The data set summaries are human generated summaries which are called reference summaries. The main purpose of this DUC is to evaluate the automatic summarization

[6] http://duc.nist.gov/

techniques, by comparing the reference summary with automatic generated summary. DUC has different tasks and reference summaries are generated within a size limits. For task one, human must generate very short summary not longer than 75 bytes. This summary is called headline summary (like newspaper headline). But in task two, human must generate short summary not longer than 665 bytes (Lin, 2004; Moghadas et al., 2013).

The data sets contain news articles from New York Times newswire. It has different subject such as computer, health care, terrorism and political issues. To build our model, 34 sets have been chosen from DUC which contain short summary with its original text.

To evaluate our model, we choose other data sets and selected 14 sets which contain original text and condensed text. Like first data sets, these sets are articles from the Economist newspaper have different subjects such as medical science, teaching and social science. The condensed texts are automatic generated summary using lexical chains (Barzilay and Elhadad, 1999).

The original texts came with different sizes such as 3KB, 5KB, 8KB, 13KB ….20KB, but the condensed texts always come with fixed size 1 KB.

For example file named APW19981019.0098 (T1) original text size is 6KB and condensed text 1KB, where file named APW19981022.0269 (T2) original text size is 4KB and condensed text 1KB and so on.

Figure 3.2 illustrate a sample of original text and figure 3.3 illustrate a sample of condensed text from DUC:

<DOC>

<DOCNO> APW19981019.0098</DOCNO>

<DOCTYPE> NEWS </DOCTYPE>

<TXTTYPE> NEWSWIRE </TXTTYPE>

<TEXT>

Britain has defense its arrest of Gen. Augusto Pinochet, with one lawmaker saying that Chile's claim that the former Chilean dictator has diplomatic immunity is ridiculous. Chilean officials, meanwhile, issued strong protests and sent a delegation to London on Sunday to argue for Pinochet's release. The former strongman's son vowed to hire top attorneys to defend his 82-year-old father, who ruled Chile with an iron fist for 17 years. British police arrested Pinochet in his bed Friday at a private London hospital in response to a request from Spain, which wants to question Pinochet about allegations of murder during the decade after he seized power in 1973. Pinochet had gone to the hospital to have a back operation Oct. 9. ``The idea that such a brutal dictator as Pinochet should be claiming diplomatic immunity I think for most people in this country would be pretty gut-wrenching stuff,'' Trade Secretary Peter Mandelson said in a British Broadcasting Corp. television interview Sunday. Home Office Minister Alun Michael acknowledged Sunday that Pinochet entered Britain on a diplomatic passport, but said, ``that does not necessarily convey diplomatic immunity.'' The Foreign Office said only government officials visiting on official business and accredited diplomats have immunity.

.

.

.

.

.

</TEXT>

</DOC>

Figure 3.2: Original Text Sample from DUC

Former Chilean dictator Augusto Pinochet has been arrested in London at the request of the

Former Chilean dictator Augusto Pinochet has been arrested in London at the request of the Spanish government. Pinochet, in London for back surgery, was arrested in his hospital room. Spain is seeking extradition of Pinochet from London to Spain to face charges of murder in the deaths of Spanish citizens in Chile under Pinochet's rule in the 1970s and 80s.The arrest raised confusion in the international community as the legality of the move is debated. Pinochet supporters say that Pinochet's arrest is illegal, claiming he has diplomatic immunity. The final outcome of the extradition request lies with the Spanish courts.

Figure 3.3: Condensed Text Sample from DUC

3.3 Concept Extraction

After choosing original texts and condensed texts, the next step is extracting concepts from both texts. This is done using ontological tool KAON[7]. Each dataset is converted to text file then uploaded to KOAN program. For example the first file named APW19981019.0098 (T1), after upload the original text and extract concepts from it; the result was 135 concepts as for this file. Table 3.1 lists samples of these concepts. For more details table, *please see table Appendix 2.*

Table 3.1: Sample of Original Text Extracted Concepts

No.	Concept	No.	Concept	No.	Concept	No.	Concept	No.	Concept
1	Defense	**28**	prosecution	**55**	Order	**82**	Guard	**109**	Television
2	Minister	**29**	Institution	**56**	Advice	**83**	Stuff	**110**	Interview
3	Priest	**30**	Power	**57**	Government	**84**	Idea	**111**	Division
4	Magistrate	**31**	Abuse	**58**	Lawmaker	**85**	Business	**112**	Riot
5	Murder	**32**	Constitution	**59**	Bed	**86**	Army	**113**	Police
21	. .		. .		. .		. .		. .
22	. .		. .		. .		. .		. .
23	Boss	**50**	Caption	**77**	Diplomat	**104**	Demonstrator	**131**	Demonstration
24	Government	**51**	Capital	**78**	Policy	**105**	Operation	**132**	Anonymity
25	Official	**52**	Publicity	**79**	Trial	**106**	Official	**133**	Reign
26	Post	**53**	Patient	**80**	Black	**107**	Event	**134**	Hospital
27	Crime	**54**	Arrest	**81**	Police	**108**	Dissident	**135**	Protest

[7] http://kaon2.semanticweb.org/

The extracting step has been done also on condensed text. The following table 3.2 shows the 20 extracted concepts.

Table 3.2: Sample of Condensed Text Extracted Concepts

No.	Concept	No.	Concept
1	Dictator	**13**	Outcome
2	Confusion	**14**	Request
3	Hospital	**15**	Charge
4	Arrest	**16**	Community
5	Murder	**17**	Surgery
6	Citizen	**18**	Death
7	Rule	**19**	Court
8	Legality	**20**	Government
9	Move		
10	Supporter		
11	Immunity		
12	Extradition		

This step has been repeated for all files. We have 34 files form the first data sets for both original text and condensed text, and 14 file form the second data sets.

Table 3.3 shows the number of extracted concepts for the first data sets. Table 3.4 shows the number of concepts for second data sets.

Table 3.3: Number of Extracted concepts of Original and Condensed Text from first Data Sets

No	Original Text File Name	Number of Original Concepts	Condensed Text File Name	Number of Condensed Concepts	Ratio between extracted concepts
1	APW19981019.009	135	D30003.M.100.T.C	20	15%
2	APW19981022.026	87	D30001.M.100.T.D	21	24%
3	APW19981030.079	131	D31022.M.100.T.C	21	16%
4	APW19981120.029	132	D30047.M.100.T.C	21	16%
5	APW19981202.127	101	D30022.M.100.T.D	16	16%
6	APW19981211.127	112	D30038.M.100.T.A	17	15%
7	APW19981212.016	184	D30053.M.100.T.B	20	11%
8	APW19981227.087	81	D30029.M.100.T.A	17	21%
9	NYT19981001.0379	397	D30027.M.100.T.C	24	6%
10	NYT19981003.0120	158	D30011.M.100.T.A	24	15%
11	NYT19981004.0102	193	D30015.M.100.T.A	19	10%
12	NYT19981010.0149	184	D30050.M.100.T.D	24	13%
13	NYT19981012.0334	221	D30036.M.100.T.D	16	7%
14	NYT19981012.0359	189	D31008.M.100.T.D	25	13%
15	NYT19981013.0354	176	D30006.M.100.T.C	26	15%
16	NYT19981013.0399	161	D31031.M.100.T.C	24	15%
17	NYT19981017.0027	127	D31026.M.100.T.C	19	15%
18	NYT19981018.0091	185	D31033.M.100.T.D	21	11%
19	NYT19981024.0050	208	D30048.M.100.T.C	22	11%
20	NYT19981104.0545	212	D30024.M.100.T.D	14	7%
21	NYT19981105.0538	194	D30008.M.100.T.D	18	9%
22	NYT19981107.0056	170	D31001.M.100.T.D	18	11%
23	NYT19981107.0251	93	D30010.M.100.T.D	17	18%
24	NYT19981114.0079	268	D30051.M.100.T.D	20	7%
25	NYT19981114.0129	185	D31013.M.100.T.D	19	10%
26	NYT19981121.0117	151	D30049.M.100.T.D	18	12%
27	NYT19981122.0163	194	D30026.M.100.T.D	25	13%
28	NYT19981126.0192	185	D30045.M.100.T.C	19	10%
29	NYT19981201.0444	155	D30005.M.100.T.C	17	11%
30	NYT19981204.0365	200	D30031.M.100.T.D	23	12%
31	NYT19981209.0451	268	D30017.M.100.T.A	23	9%
32	NYT19981219.0117	235	D30046.M.100.T.C	16	7%
33	NYT19981221.0377	178	D31050.M.100.T.E	23	13%
34	NYT19981223.0347	229	D30033.M.100.T.D	22	10%

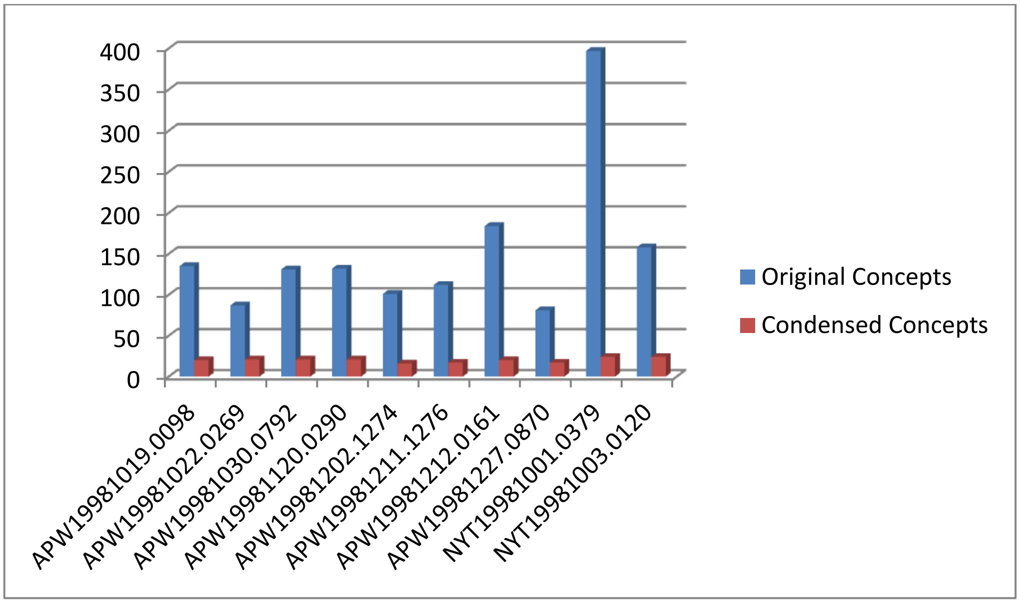


Figure 3.4: Sample of Extracted Concepts Percentage for First Data Set

Figure 3.4 shows an example of the ratio between the number of extracted concepts from the original text and number of extracted concepts from the condensed text. Data set summary size is less than 1KB (665 byte) and the original text size comes from 5KB to 20KB. From table 3.3, we can note the maximum ratio between the extracted concepts from the original text and the extracted concepts from the condensed text is 24% and the minimum is 6%. This is because of the variance between the size of the original text and the condensed text. Also when we extracted the concepts using KOAN, we used the default frequency of concepts which was (3). These two conditions have been applied on all data sets.

Table 3.4: Number of Extracted concepts of Original and Condensed Text from Second Data Sets

No.	Original Text File Name	Number of Original Concepts	Condensed Text File Name	Number of Condensed Concepts	Ratio between extracted concepts
1	Text 1	114	Summary 1	23	20%
2	Text 2	228	Summary 2	30	13%
3	Text 3	91	Summary 3	16	18%
4	Text 4	89	Summary 4	15	17%
5	Text 5	53	Summary 5	15	28%
6	Text 6	87	Summary 6	9	10%
7	Text 7	127	Summary 7	24	19%
8	Text 8	86	Summary 8	14	16%
9	Text 9	79	Summary 9	23	29%
10	Text 10	121	Summary 10	15	12%
11	Text 11	176	Summary 11	40	23%
12	Text 12	335	Summary 12	40	12%
13	Text 13	60	Summary 13	10	17%
14	Text 14	88	Summary 14	32	32%

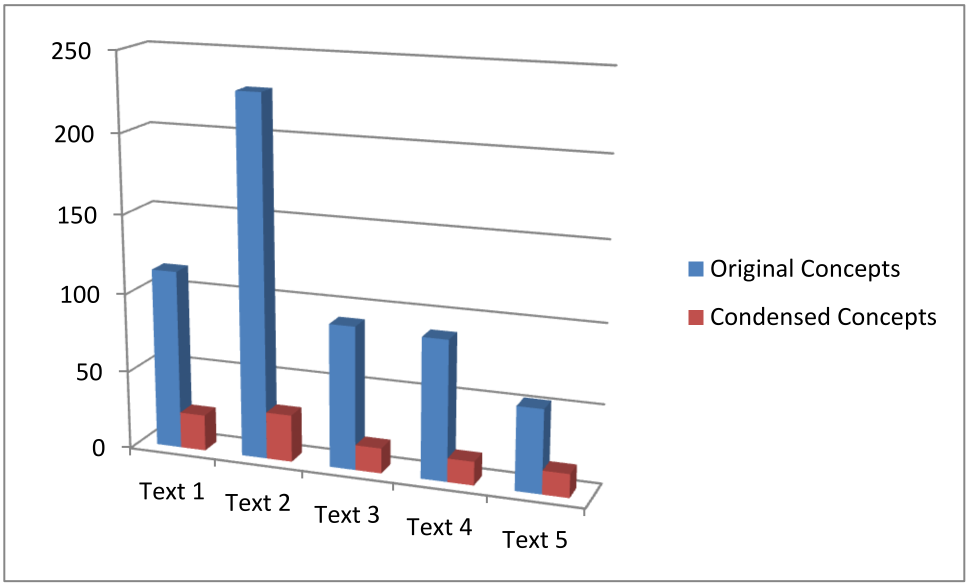


Figure 3.5: Sample of Extracted Concepts Percentage for Second Data Set

Figure 3.5 shows an example of ratio between number of extracted concepts from the original text and number of extracted concepts from the condensed text in second data set. From table 3.4, we can note the maximum ratio between the extracted concepts from the original text and the extracted concepts from the condensed text is 32% and the minimum is 10%, the average ratio is around 19%. When we extracted the concepts using KOAN, we used the default frequency of concepts which is (3). Also the condensed text size comes with fixed size 1KB, therefore; the condensed text size is the same for all data set.

3.4 Applying the Measures

This step took long time as we need to apply it for 34 file each file must apply six different measures. This is done using WS4J Demo software. The following part discusses in details how we apply the measures. We choose file name (T1) as an example to explain our experiments.

3.4.1 Upload Concepts in WS4J

After extracting the concepts from each original text and condensed text, each file contains a number of concepts. For example, the first file (T1) contains 135 concepts from original text and 20 concepts from condensed text. WS4J[8] has two inputs options, we choose to match a set of concepts at once because we got large number of concepts and this will save time and effort. Figure 3.4 shows how this is done. To calculate the semantic similarity for all the extracted concepts, we need to upload the 20 concepts and compare them with all the 135 concepts from the original text.

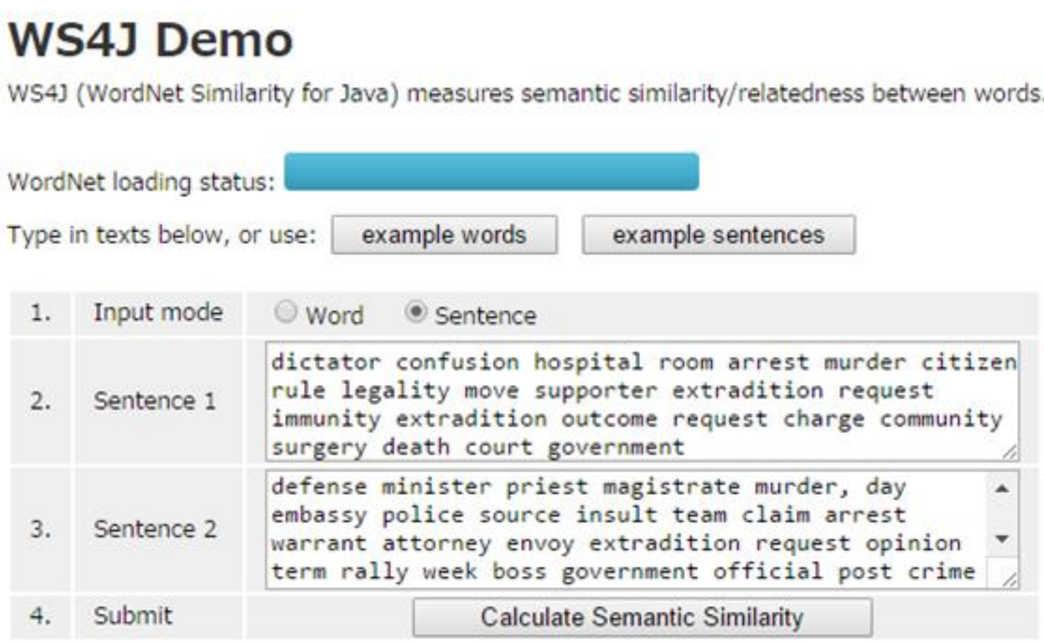


Figure 3.6: Calculate Concepts Semantic Matching

---

[8] http://ws4jdemo.appspot.com/

3.4.2 Calculate the Results for all Measures

After collecting data sets that contain both original text and condensed text. Then extract the concepts from each original text and condensed text. The next step is to match between the extracted concepts using different semantic measures. Table 3.5 clarifies a sample of using WuP measure for the first file T1. Columns represent the number of extracted Original Concepts (OC), where the table's rows represent the number of extracted Condensed Concepts (CC). Each file has six Semitic matching results using (WuP, LCH, LIN, Resnik, HSO, and LESK) measures. Each file has six tables. *For full table refer Appendix3.*

Table 3.5: Sample of Semantic Matching for WuP Results

OC	CC1	CC2	CC3	CC4	………	CC17	CC18	CC19	CC20
	dictator	Confusion	hospital	Arrest		Surgery	death	Court	Government
Defense	0.5556	0.6667	0.7778	0.6667	………	0.7778	0.625	0.8235	0.8
Minister	0.7273	0.6	0.4545	0.6	………	0.6667	0.5263	0.6667	0.6
Priest	0.7619	0.6667	0.4762	0.6667	………	0.4762	0.7143	0.6957	0.375
Magistrate	-	-	-	-	……	-	-	-	-
Murder	0.2	0.6667	0.2727	0.5455	………	0.5217	0.88	0.6087	0.5455
Day	0.7826	0.625	0.4762	0.625	………	0.4762	0.6667	0.6957	0.4286
Embassy	-	-	-	-	………	-	-	-	-
Police	0.2667	0.4	0.7059	0.4	………	0.3333	0.4286	0.7059	0.75
Source	0.8182	0.4444	0.6667	0.4444	………	0.6667	0.5714	0.7273	0.6
Insult	0.2857	0.6316	0.375	0.6316	………	0.7	0.5556	0.6	0.6316
Team	0.2857	0.4286	0.7059	0.4286	………	0.3529	0.4615	0.7059	0.75

For example : the two concept matching from the (T1) file , first condensed concept CC1 is *dictator* and first original concept OC1 is *defense*, the result as shown in the table 3.6 using WuP measure is: 0.5556.

Table 3.6: Two Concepts Matching Result

OC	CC1
	Dictator
Defense	0.5556

3.5 Evaluation

The evaluation process is to evaluate the quality of the condensed text by comparing it with the original text. Evaluation process first done by human then by comparing human evaluation and semantic measures evaluation then by calculating the errors of each measure and checking which measure that gives the minimum error. We will use descriptive statistical Mean Square Error (MSE) to evaluate the final results.

CHAPTER FOUR

Matching Process and Experiment

Overview

This chapter explains in details the proposed model. This is done after collecting the data sets and extracting main concepts from both condensed text and original text. The aim is to find which measure from semantic measures with a minimum error. The experiments are divided into two types of evaluations. In the first evaluation, the selected condensed texts have covered the original texts with 100%. In the second evaluation, expert evaluated the condensed text coverage for the original texts.

4.1 Proposed Model

Our proposed model contains the following phases:

1. Extracting concepts from the original text.
2. Extracting concepts from the condensed text.
3. Computing semantic measures among concepts.
4. Finding the best semantic measure with a minimum error.

The first and the second phase have been explained in chapter three. The following section will explain in details the semantic matching phase.

This research collected two data sets. For each set, we extracted the concepts from both original and condensed texts. In this phase, we are going to compare each concept from the original text with the concept from the condensed text. For example in file (T1), the extracted concepts from the original text were 135 concepts, and from the condensed text were 20 concepts. Table 4.1 shows an example of applying WuP measure on the 135 concepts against the 20 concepts for the original and condensed text respectively. The first field of the table 4.1 represents the Original Concept (OC); in this example the concept is "defense". The first row of the table 4.1 also represents the Condensed Concepts (CC) that has been extracted from the condensed text. For each concept from the 135 concepts we computed the WuP measure value. Table 4.1 represents a sample for this WuP measure. *For full results see appendix number 3*. This step has been repeated for all the files and for the six measures (34 files from first data sets and 14 files from second data sets).

Table 4.1: Samples of Matching Original text Concepts with Condensed Concepts using WuP

OC	CC1	CC2	CC3	CC4		CC17	CC18	CC19	CC20
	dictator	**Confusion**	**Hospital**	**arrest**	…………	**surgery**	**death**	**Court**	**government**
defense	0.5556	0.6667	0.7778	0.6667	…………	0.7778	0.625	0.8235	0.8
minister	0.7273	0.6	0.4545	0.6	…………	0.6667	0.5263	0.6667	0.6
priest	0.7619	0.6667	0.4762	0.6667	…………	0.4762	0.7143	0.6957	0.375

Applying six semantic measures for 135*20 concepts for 5 acceptance rates produced more than 80,000 data items for only one file (135*20*6*5=81,000).

4.2 Calculate the Maximum for Each Concepts

The maximum is important to see how far these concepts are closed to each other. At first, we need to define the maximum value for each semantic measure (SM). Some semantic measures have maximum value such as WuP and LIN where they have "1" as a maximum value while the maximum value of HSO measure is “16”. The other measures have no maximum value. The aim is to find how the condensed concepts best covered the original text concepts at a certain point. For example in the file (T1), the goal is to find how much the 20 condensed concepts cover the 135 original concepts. This is mean which concept from the 20 concepts that have the maximum matching value, and we can accept only if the maximum go beyond a certain point (cutting point). As table 4.2 shows that the maximum value is between original concepts “defense” and condensed concepts “charge” which has the value (0.8889) using WuP measure.

Table 4.2: Calculate Maximum Value Example using WuP

No	OC	CC1	CC2	....	CC15	CC16	CC17	CC18	CC19	CC20	MAX
		Dictator	confusion	......	charge	community	surgery	death	Court	Government	
1	defense	0.5556	0.6667	....	0.8889	0.75	0.7778	0.625	0.8235	0.8	0.8889

4.3 Calculate the Maximum for Each Semantic Measure

As we discussed chapter three, semantic measures can be divided into semantic similarity measures and semantic relatedness measures. Some measures has maximum value (WuP, LIN, HSO) while other has no maximum value (LESK, LCH, Resnik). Measures such as WuP & LIN have a maximum value which is greater than or equal to zero and less than or equal to one. HSO measure has maximum value which is greater than or equal to zero and less than or equal to 16. We need the maximum value to calculate the value for each cutting points. Cutting point means that if semantic measures between two concepts above this point; we will accept it otherwise if it’s below this point, we reject it. There is a problem to define this; therefore, the maximum will help us to know if the result below this certain point the result is rejected. For example using WuP measure, the maximum value for this measure is “1”. To calculate the number of accepted concepts for a certain point for example “50%”. This means all the semantic matching results that have the maximum value above “0.50” is accepted, other matching results which is below “0.50” is rejected.

In the rest three measures (LESK, LCH, Resnik) there is no maximum value. Therefore, we cannot calculate the percentage of cutting point of each file. Thus, four different techniques are proposed to calculate the maximum number of each file. Then we can calculate the cutting point with a different percentage. Next part illustrates an explanation for each technique; we choose LESK measure to illustrate the techniques:

1) Max average:

We took the average of the maximum value for each original concepts (OC) and condensed concepts (CC).

For example file (T1), the original text has 135 extracted concepts and the condensed text has 20 extracted concepts. We compute the semantic matching for each OC with the CC. Then we calculated the maximum results. Thus, we will have 135 maximum for this file. After calculating the semantic measures results, we choose LESK measure to illustrate the example. The average value for all 135 maximums was 645 (LESK measure).

2) Trimming Max Average with 5%

We used the trimming method with 5% after sorting the semantic measure maximum from low to high.

For example in file (T1), the extracted concepts form the original file is 135 concepts and the extracted concepts form the condensed file is 20 concepts. We computed the semantic measures matching for all the extracted concepts, and then calculated the maximum results. The maximum values are sorted in ascending. 5% trimming from 135 is "7" concepts. Therefore, we removed 7 concepts form below and 7 concepts from above. Then, we calculate average of the remaining concepts "121 concepts" after trimming and the average was 300 (LESK measure).

3) Average for all results

This is done by calculating the average of semantic matching results for all the extracted concepts from the original text OC, and the extracted concepts from the condensed text CC. We need to calculate the result as number of OC * number of CC for a given file.

For example in file name (T1) which we discussed before that it's contained 135 OC and 20 CC. So the result will be the average of (135*20) = 2700 and the value was 98 (LESK measure).

4) Trimming Average for all results with 5%

This technique is close to number two, but instead we sorted the concepts from high to low. Then use trimming method with 5% percentage and chooses the first max after trimming.

For example in file name (T1) as explained before, the extracted concepts form the original file is 135 concepts and the extracted concepts form the condensed file is 20 concepts. The results from matching are 135 maximum values. The maximum values are sorted in descending way. 5% trimming from 135 is “7” concepts. Therefore, we removed 7 concepts from the above. The result of the first maximum value after trimming was 2595(LESK measure).

In all fourth methods, we got four values (645, 300, 98, and 2595). We need to choose neither not very high number nor small one because if we took the low value “98” most of the concepts will be included in our experiment. At the same time, if we took high value “2595” most of the concepts will be ignored and not included. So we decided to choose average value which is 300 as the maximum value for this file.

These techniques are tested for three measures (LESK, LCH, and Resnik). We found that the best technique is to calculate maximum is technique number two, which is trimming Max Average with 5%.

The Trimming Max Average with 5% is used for both first data sets and second data sets and for the three measures which we discussed before. Each set has different extracted original concepts. Therefore, the trimming with 5% value is depends on the number of the original text extracted concepts.

4.4 Semantic Matching with Different Cutting Points

This part explains why the researcher chooses different cutting points. Also discuss the result for each one. To accept how each two concepts are closed or far from each other; we need a cutting point. To find which cutting point is the best is critical. Therefore, we choose five different points with different values (40%, 50%, 60%, 70%, and 80%) to accept or reject the matching results, and then apply them to all semantic measures. For each original text and it's condensed we measure these cutting points for the six semantic measures. The equation (6) shows our own equation to calculate the acceptance rate for a certain measure. Therefore, in data set one we apply this for 34 files and for the six measures. Same done in data set two which has 14 files.

$$\textbf{Acceptance Rate \%} = \frac{\textbf{No. of concepts above cutting point \%}}{\textbf{No. of original Text concepts}}\ \mathbf{100\%} \quad \ldots\ldots (6)$$

Different cutting points are used to check the coverage of each condensed text. Each measure has five different cutting points. Next section illustrates an example for some of the semantic measure with different cutting points.

4.4.1 Examples of Different Acceptance Rate

Cutting point percentage calculation based on how the condensed concepts cover original concepts with threshold value. For example, first file (T1) form first data set; the result of cutting point 40% for WuP measure is 90.37. This is by calculating 40% from the maximum value of WuP measure. As discussed before, the maximum value for WuP is 1. Thus, for 40% the value is 0.40. That's mean all concepts with maximum is equal or greater than 0.40 is counted.

In the same file using cutting point 40% with WuP measure. This part illustrates how the calculation is done. The aim is to find how the condensed text covers the original text. The number of extracted concepts form condensed text is 20 concepts and number of extracted concepts from original text is 135. Therefore, we need to find how much these 20 CC covered the OC which is 135. The researcher chooses different point to accept or reject the results. The total number of concepts which has a maximum value equal or greater than 0.40 is 122 concepts; this mean that the 20 condensed concepts is covered with 122 original concepts. To calculate the coverage percentage, we divided the covered concepts 122 by the total number of original text extracted concepts which are 135 concepts. The result using equation (6) is 90.37 as showing in table 4.3 for the first file using WuP measure. For the second file (T2), the result is 87.36.

$$\textbf{Acceptance Rate in 40\% (T1)} = \frac{\mathbf{122}}{\mathbf{135}} * \mathbf{100\%} = \mathbf{90.37}$$

Table 4.3: Sample of Result for Cutting Point 40% in WuP Measure

File Name	Text	OT size	CT size	40% cutting point
APW19981019.0098	T1	8 KB	1 KB	90.37
APW19981022.0269	T2	4KB	1 KB	87.36
APW19981030.0792	T3	5KB	1KB	86.26
APW19981120.0290	T4	5KB	1KB	87.88
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…
NYT19981223.0347	T34	11KB	1KB	93.89

Table 4.3 presents a list of examples for data set one. Its show each file name and original text size" OT size" and condensed text size "CT size". It also shows sample 40% cutting point result using WuP. We will discuss later how to calculate the average error for each cutting point.

Another walk through example for the first file (T1) but using another cutting point which is 70% using HSO measure. As explained previously, the maximum value for HSO is 16. Thus, we need to measure the degree of acceptance and rejection results to be above 70% cutting point, 70% from 16 is equal 11. This means all concepts with maximum are equal or greater than 11 are counted. The total number of concepts which has a maximum value above 11 is 20 concepts. This means that the 20 condensed concepts covered 20 original concepts. The result is 14.81 as showing in table 4.4 for the first file using HSO measure.

$$\text{Acceptance Rate in } 70\% \text{ (T1)} = \frac{20}{135} * 100\% = 14.81$$

Table 4.4: Sample of Result for Cutting Point 70% in HSO Measure

File Name	Text	OT size	CT size	70% cutting point
APW19981019.0098	T1	8 KB	1 KB	14.81
APW19981022.0269	T2	4KB	1 KB	21.84
APW19981030.0792	T3	5KB	1KB	14.50
APW19981120.0290	T4	5KB	1KB	16.67
…	…	…	…	…
…	…	…	…	…
…	…	…	…	…
NYT19981223.0347	T34	11KB	1KB	3.93

4.4.2 Result for Different Cutting Point

This part shows why the researcher chooses different cutting point. For each file, six different measures are applied in the condensed text with five different cutting points. This mean that the experiments for the first data set which contain 34 set is (34*6*5=1020 experiments). For second data set which is 14 files (14*6*5=420 experiments).

Table 4.5 shows sample of results using different cutting points in WuP measure. By choosing high value cutting point such as 80%, this means that we need to high coverage between condensed text and original text. From table 4.5, we can notice that the lower cutting point value the higher result. For example first file (T1) using 40% cutting point; this mean all the maximum value that is above 0.40 is counted. The result was 90.37 (122 concepts that have value above 0.40 divided by the total number of original extracted concepts which is 135). While the result at the same file (T1) but using different cutting point 80% is 64.44 (87 concepts that have value above 0.8 divided by the total number of original extracted concepts135).

Table 4.5: Sample of Result for all Cutting point in WuP Measure

File Name	Text	40% cutting point	50% cutting point	60% cutting point	70% cutting point	80% cutting point
APW19981019.0098	T1	90.37	90.37	90.37	87.41	64.44
APW19981022.0269	T2	87.36	87.36	86.21	80.46	60.92
APW19981030.0792	T3	86.26	86.26	83.97	79.39	62.60
APW19981120.0290	T4	87.88	87.12	84.85	79.55	66.67
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
NYT19981223.0347	T34	93.89	93.89	93.01	88.21	68.56

4.5 Two Cases

This research classifies the experiments in two parts. The classification is based on human evaluation of how much the condensed text covers its original text. First part, we consider the data sets' condensed text as an ideal, so we assume that the condensed text has covered the original text and all condensed text has 100% coverage. Second part, we considered that this condensed text does not cover the original text and we need expert to evaluate the coverage of the condensed text. Expert evaluation divided into three categories (Good coverage"70% - 90%", Medium coverage "60% - 75%, and Bad coverage "40% - 55 %"). Next section will discuss in details each experiment.

4.5.1 Ideal Summaries

This part we considered the data sets' condensed text as an ideal and no need for expert to evaluate it. Thus, we consider all human evaluation as 100 %. This mean that all the condensed text is covered the original text in data sets one. The aim is to explore which measure from the semantic measures that can give the minimum error. Through the results, we can find the best measure that can give minimum error in case we are sure that the condensed text is covered the original text. This is done only in data set one, as it's considered as a reference or an ideal summary.

Table 4.6 shows sample of some results using WuP measure. It also represent human evaluation field. Through the table we can notice that human evaluation for all files is 100%. This is because we considered the summary as an ideal summary for each original text. Thus, we assume that all files has 100% coverage between the condensed text covered the original text. After that, we need to calculate the error between human and each cutting point. The error calculation is the difference between human and a given cutting point. The calculation as below:

Error for a cutting point = (Human Evaluation - Cutting point result) ^{2}…….. (7)

For example in first file (T1), the result using WuP measure in cutting point 40% is 90.37. To calculate the error for cutting point 40% , the results is the difference between human evaluation which is 100 here and the result for error in 40% cutting point which is here 90.37. Thus the result for the error using WuP measure in 40% cutting point is:

Error cutting point 40% = $(100 - 90.37)^2 = 93$

To be able to compute the mean square error, we record the error between the human evaluation and each cutting point result. This is done for different cutting points (40%, 50%, 60%, 70%, and 80%) and for each semantic measure (WuP, LCH, LIN, Resnik, HSO, and LESK). More details *refer to appendix 4.*

Table 4.6: Result for Cutting Point 40% in WuP with Human Evaluation 100%

File Name	Text	Human Ev. %	40% cutting point	ERROR 40%
APW19981019.0098	T1	100	90.37	93
APW19981022.0269	T2	100	87.36	160
APW19981030.0792	T3	100	86.26	189
APW19981120.0290	T4	100	87.88	147
…	…	100	…	…
…	…	100	…	…
…	…	100	…	…
NYT19981223.0347	T34	100	93.89	37

4.5.2 Different Evaluation Rates

In this part, an expert evaluation is needed to find whether if the condensed text covers its original text. We considered that condensed text does not cover the original text. That’s why; an expert evaluation is needed to evaluate how much the condensed text cover the original text. Expert evaluation divided into three categories:

Good coverage: 70% - 90%.

Medium coverage: 60% - 75%.

Bad coverage: 40% - 55%.

Original text files come with different sizes such as 4KB, 9KB, 15 KB and 20KB. But the condensed texts come with fixed size 1 KB. Because of that, an expert evaluation is important to find if the condensed text cover most content of the original text. By Good coverage we mean that the condensed text covers most of the content of its original text. In the other hand, bad coverage means that the condensed text does not cover the original text.

Table 4.7: Expert Evaluation Percentage for First Data Set

File Name	Text	OT size	CT size	Human Ev. %
APW19981019.0098	T1	8 KB	1 KB	70
APW19981022.0269	T2	4KB	1 KB	85
NYT19981114.0129	T25	15KB	1KB	55
APW19981120.0290	T4	5KB	1KB	80
NYT19981004.0102	T11	9KB	1KB	65
NYT19981001.0379	T9	20KB	1KB	40
…	…	…	…	…
…	…	…	…	…
NYT19981223.0347	T34	11KB	1KB	65

Table 4.7 present sample of expert evaluation percentage for first data set. It also shows each original text size (OT) and condensed text size (CT). For example first file (T1) with

8KB original text size and 1KB the condensed text size. The expert evaluation was 70%; this means when human compare the coverage between condensed text and original text, the result is the coverage is medium with 70%. While file (T9) the expert find the coverage of the condensed text is 40%, this means that the condensed text has a bad coverage comparing it with its original text.

The expert evaluation part is done for all 34 files from the first data sets and 14 file for the second data sets. In this experiment we need to find which measure can give a result that is closed to expert evaluation. And also which measure that can give the minimum. This experiment also needs different cutting point to compute the coverage of the condensed text.

Table 4.8: Result for Cutting Point 40% in WuP with Expert Evaluation Rates

File Name	Text	OT size	CT size	Human Ev. %	40% cutting point	ERROR 40%
APW19981019.0098	T1	8 KB	1 KB	70	90.37	415
APW19981022.0269	T2	4KB	1 KB	85	87.36	6
APW19981030.0792	T3	5KB	1KB	85	86.26	2
APW19981120.0290	T4	5KB	1KB	80	87.88	62
…	…	…	…	…	…	…
…	…	…	…	…	…	…
…	…	…	…	…	…	…
NYT19981223.0347	T34	11KB	1KB	65	93.89	834

Table 4.8 shows Result for cutting point 40% in WuP with Human Evaluation rate. To calculate the error for each cutting point we need to find the difference between human evaluation and the result for each cutting point. For example first file (T1); the expert evaluated the coverage of the condensed text by comparing with its original text. The

expert evaluation result was 70%. As discussed before, the calculation using WuP measure in cutting point 40% is 90.37. Therefore, the error between expert evaluation result and cutting point 40% result is the difference between both. The result is:

Error in 40% = (70 - 90.37)² = 415

This calculation is the same as has been shown in the next file where the error is (6). This is done for all cutting points. For WuP measure, it has five different cutting points. Same is done for the other semantic measures. Through the results of these experiments, we need the find which measure from the six measures can give minimum error in a certain cutting point.

4.6 Experiment Result and Analysis

This research used Mean Square Error (MSE) to calculate the average error of each semantic measure. Each measure has different five cutting point, so each measure has five MSE result (Chai & Draxler, 2014).

$$\textbf{MSE} = \sum_{n}^{1} \sqrt{\frac{(\textbf{Human Evaluation} - \textbf{Cutting point Result})^2}{\boldsymbol{n}}} * 100\%$$

As discussed before, this research considered that the condensed texts cover its original text. Thus, in this part the expert evaluation for this part is 100%. The other part of experiment assumes that the condensed text does not cover the original text. Therefore, expert evaluation is needed. Results for expert evaluation come with different percentage based on the coverage of the main content in condensed text. Next section explains the result for both parts.

Result for Ideal Summaries

As explained before, this research used six semantic measures (WuP, LCH, LIN, Resnik, HSO, and LESK) and five different cutting points (40%, 50%, 60%, 70%, and 80%). Each measure had five different results based on cutting point percentage. After we calculate the difference between expert evaluation and each measure result; Next step is to compute the average error for each measure in a given cutting point. Figure 4.1 present average errors for cutting point 40% for the six measures. Through the table we can note that WuP measure and LCH measure have lowest error with (12.22, 11.80). From results, we realized that measure that can give the minimum average error is LCH with (11.80) error. HSO measure gives the maximum error with (78.45). For the rest figures of all cutting points error, *refer to appendix no. 5.*

Table 4.9: Average MSE for 40% Cutting Point

Measures	Cutting point 40%
WuP	12.22
LCH	**11.80**
LIN	31.86
Resnik	18.38
HSO	78.45
LESK	42.27

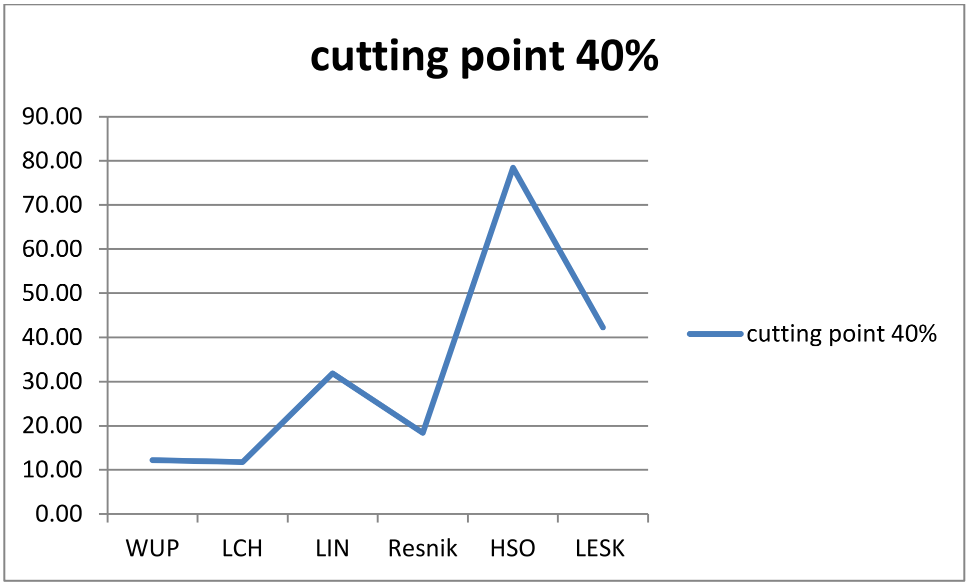


Figure 4.1: Average MSE for the 40% Cutting Point

Through table 4.10, we concluded that LCH measure is the best measure for cutting point 40%. This means if we have a condensed text and this condensed text has a high coverage of its original text; we can use LCH measure. We can note that LCH has the minimum error in most cases, also the average error for LCH between cutting point 40% (11.80) and cutting point 50 %(11.83) is insignificant.

Table 4.10: Average MSE for all Cutting Points

Measures	cutting point 40%	cutting point 50%	cutting point 60%	cutting point 70%	cutting point 80%
WuP	12.22	12.67	15.28	20.75	38.41
LCH	**11.80**	11.83	12.48	15.79	26.04
LIN	31.86	44.57	57.25	69.06	77.68
Resnik	18.38	20.46	25.46	31.79	40.31
HSO	78.45	87.98	87.98	87.98	87.98
LESK	42.27	50.10	56.15	61.51	66.14

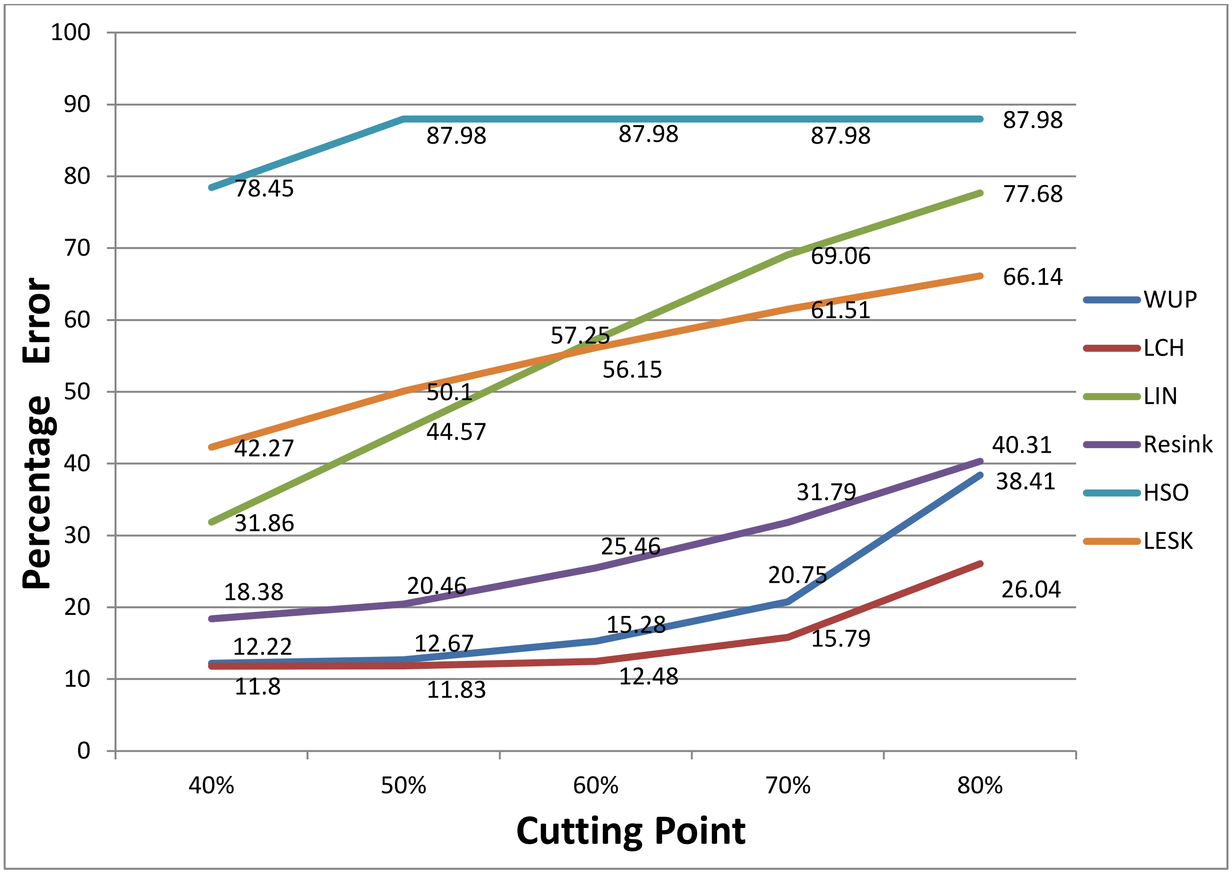


Figure 4.2: Average MSE for all Semantic Measures

Figure 4.2, shows that LCH measure is the best measure with minimum error. The figure also shows that there is a crossing point between LESK measure and LIN measure for the 60%cutting point. It's clear that LIN measure has error less than LESK. For example, for cutting point 40% LIN has 31.86 where LESK has 42.27. From that we concluded if we are looking for a generalized summary, cutting point is less than 60%, LIN gives minimum error comparing with LESK. However, if we are looking for a specific summary LESK gives minimum error comparing with LIN measure.

Results for different evaluation rates

As we discussed before, this experiment used expert evaluation to evaluate the quality of the condensed text. This is done by comparing the coverage with its condensed text. To compare the result of expert evaluation and out result; first we extracted the concepts from original text and condensed text, then applying the six semantic measures on the extracted concepts. Then we calculated the average error for each semantic measure. The results show that Resnik has the minimum error with 12.21% on average. Table 4.11 and figure 4.4 presents the results at cutting point 70%.

It's clear that Resnik measure has minimum error value.

Table 4.11: Average MSE for the 70% Cutting Point

Measures	cutting point 70%
WuP	16.24
LCH	19.32
LIN	40.62
Resnik	12.21
HSO	58.36
LESK	34.19

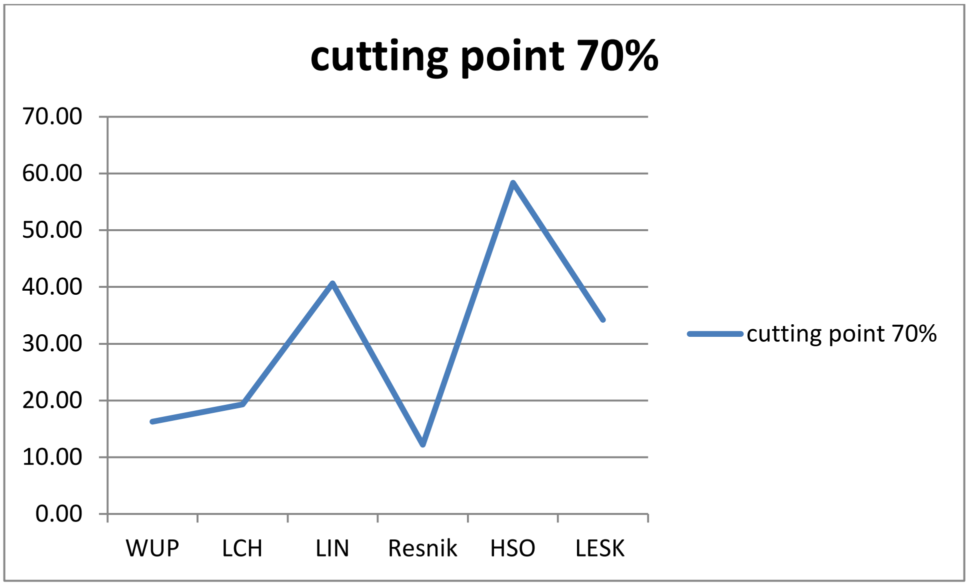


Figure 4.3: Average MSE for the 70% Cutting Point

Table 4.12: Average MSE for all Cutting Points

Measures	cutting point 40%	cutting point 50%	cutting point 60%	cutting point 70%	cutting point 80%
WuP	22.12	21.65	19.55	16.24	16.08
LCH	22.41	22.35	21.69	19.32	14.38
LIN	13.04	19.83	30.22	40.62	48.66
Resnik	17.21	16.24	13.17	**12.21**	15.96
HSO	49.23	58.36	58.36	58.36	58.36
LESK	20.24	25.26	29.72	34.19	38.39

Table 4.12 lists all results for all semantic measures using different cutting points. Through figure 4.4, we concluded that if we have a condensed text and we are not sure about the summary quality. We need an expert to evaluate it, to define whether the condensed text covers all the content of its original text.

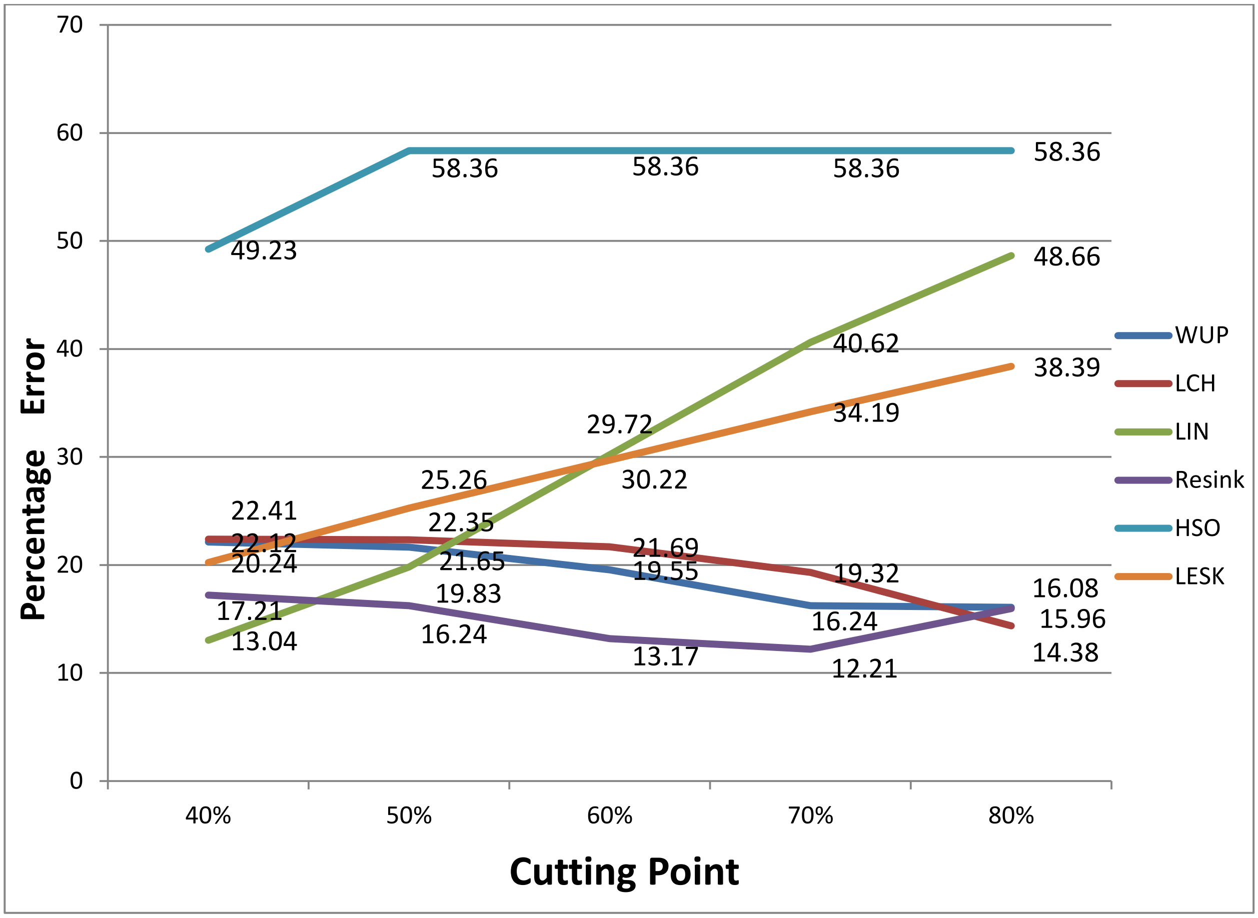


Figure 4.4: Average MSE for all Semantic Measures

From figure 4.4 shows that there are many crossing points between the measures. When the cutting point is less than 60%, i.e. the summary is very general, the error is not clear. However, when the cutting point is more than 60% the error become clear. For example, the crossing point for Resnik measure is (17.21) where LIN measure is (13.4) at cutting point 40%. This means if we are looking for a generalized summary LIN measure gives the minimum error with MSE 13.4%. However, for a precise summary, Resnik has the minimum error in most cutting points.

Another crossing point is between LCH measure with average error (14.38) and WuP with average error (16.08). LCH and WuP are from the same family which is the path length. Both depend on calculating the depth between two concepts based on the node counting. The depth in given taxonomy is the length of the shortest path for two concepts between the root of the taxonomy and these concepts.

In the case of bad summaries; i.e. when the expert give 40% coverage In this case, this research recommends using LESK measure. Table 4.13 shows the results for all bad coverage from first data sets for (T9), column (H.E) shows the results of human evaluation which was (40%). It also shows the results for different acceptance rates used with the six semantic measures. Table 4.14 shows the result for the second data sets (T12). Both tables show that LESK measure gave the minimum error. Table 4.13 shows that LESK measure has the minimum error (3%). Table 4.14 shows that LESK measure has the minimum error (1%).

The results from table 4.13 and 4.14 show that LESK is the best measure to evaluate the bad coverage summary. We concluded that the semantic measures can be used to identify not only the good coverage but also the bad coverage.

Table 4.13: Result for Bad Coverage from First Data Sets

File Name	Text	OT size	H.E	40%	Cutting Points ………	80%	40%	Error ………	80%	SM
NYT19981001.0379	T9	20KB	40	90	………	64	**2467**	………	**563**	WuP
NYT19981001.0379	T9	20KB	40	84	………	61	**1925**	………	**450**	RES
NYT19981001.0379	T9	20KB	40	72	………	24	**1011**	………	**266**	LIN
NYT19981001.0379	T9	20KB	40	64	………	38	**599**	………	**3**	LESK
NYT19981001.0379	T9	20KB	40	90	………	79	**2518**	………	**1489**	LCH
NYT19981001.0379	T9	20KB	40	22	………	10	**309**	………	**911**	HSO

Table 4.14: Result for Bad Coverage from Second Data Sets

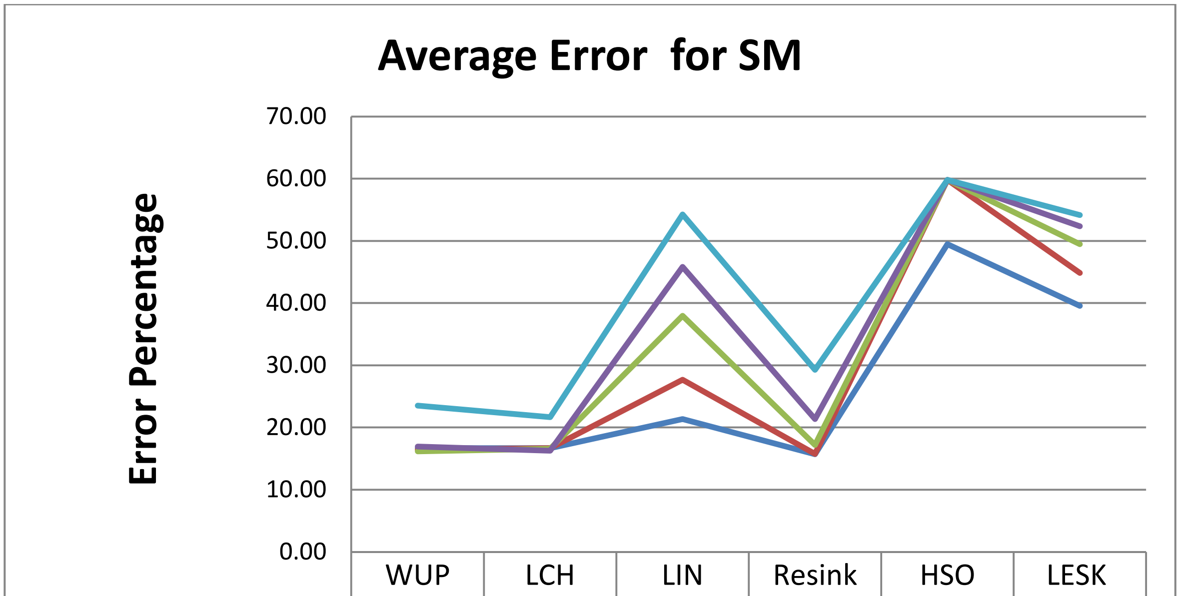
File Name	Text	OT size	H.E	…….	60%	Cutting Points 70%	80%	…….	Error 60%	70%	80%	SM
Text 12	T12	19KB	40	…….	89	85	74	…….	**2397**	**2005**	**1158**	WuP
Text 12	T12	19KB	40	…….	79	74	68	…….	**1506**	**1158**	**787**	RES
Text 12	T12	19KB	40	…….	58	44	33	…….	**321**	**13**	**56**	LIN
Text 12	T12	19KB	40	…….	41	36	32	…….	**1**	**17**	**70**	LESK
Text 12	T12	19KB	40	…….	89	86	76	…….	**2426**	**2113**	**1326**	LCH
Text 12	T12	19KB	40	…….	18	18	18	…….	**475**	**475**	**475**	HSO

4.7 Evaluate the Result using Second Data Set

This research used two data sets. The first data set contains 34 original texts and condensed texts. We apply our experiments in the first data set; but to evaluate the results, we used another data set which contains 14 original texts and 14 condensed texts. In this set, we assumed that the condensed text does not cover the original text. Therefore, an expert evaluation is needed. All steps have been repeated as for the first data set. Then the error has been calculated by computing the difference between expert evaluation and the value of each measure. Table 4.13 and figure 4.5 show the average error using different semantic measures. The results show that Resnik measure is the best measure that gives minimum error 15.68% using cutting point 40%.

Table 4.15: Average MSE for all Cutting Points using Second Data sets

Measures	cutting point 40%	cutting point 50%	cutting point 60%	cutting point 70%	cutting point 80%
WuP	16.70	16.43	16.15	16.96	23.50
LCH	16.68	16.70	16.56	16.27	21.69
LIN	21.34	27.68	37.98	45.80	54.69
Resnik	**15.68**	**15.77**	17.22	21.38	29.79
HSO	49.50	59.79	57.79	59.79	59.79
LESK	39.53	44.86	49.46	52.37	54.15



	WUP	LCH	LIN	Resink	HSO	LESK
cutting point 40%	16.70	16.68	21.34	15.68	49.50	39.53
cutting point 50%	16.43	16.70	27.68	15.77	59.79	44.86
cutting point 60%	16.15	16.56	37.98	17.22	59.79	49.46
cutting point 70%	16.96	16.27	45.80	21.38	59.79	52.37
cutting point 80%	23.50	21.69	54.25	29.27	59.79	54.15

Figure 4.5: Average MSE for all Semantic Measures in Second Data Sets

CHAPTER FIVE

Conclusions and Future Work

Overview

This chapter summarizes the work done through this research. It discusses the conclusion from the results in both experiments. It also discusses future work.

5.1 Conclusion and Contributions

The thesis concluded in finding the best measure that can give a minimum error for measuring the quality of a condensed text. The results from first experiment showed that LCH measure has the minimum error comparing with other measures. This means that if you are sure that these summaries are the best ones for a given texts, you can use LCH to define how much the quality of these condensed texts.

However, if you believe that these summaries are not good ones and you need to find the quality of these summaries. Using Resnik measure will give the minimum error comparing it with other measures. Thus, Resnik is the best measure to define whether a condensed is covered or not.

This thesis also found out the best measure that can give the bad coverage of condensed text.

This means if we have a summary and we need to find out how much this summary is bad. LESK will give the minimum error. Thus, LESK is the best measure to define the worst (bad coverage) condensed text.

The above contributions can be summarized in one main point which is the semantic measures can be used to identify not only the good coverage but also the bad coverage.

The research evaluated the semantic measures and showed how these measures could be used to evaluate the quality of a condensed text. Also this research contributed in defining the best semantic measures with minimal error that defines how much a condensed text covers its original text.

The main steps for this research could be summarized in the following points:

1) This research studied how to find the quality of the condensed text using the semantic measures.

2) In order to study the quality of the condensed text. We used data sets that contain the original text with its condensed text. We collected data sets (first data set 34 files and second data set 14 files).

3) Extraction concepts from each condensed text and original text using KAON.

4) Semantic matching using six different measures (WuP, LCH, LIN, Resnik, HSO, and LESK). Each two measures belong to one family.

5) We calculated the coverage using different cutting point for each measure. By choosing different cutting point; we aim to find how closed the condensed text concepts from original text concepts.

6) The final results show that using a measure to define the coverage of the condensed text is based on the summary quality. We have three cases: if you are looking for the ideal summary, you can use LCH measure. If you are not sure from the quality of the condensed

text, you can use Resnik. If you are looking for a summary with a bad coverage, the measure that can tell that is LESK.

5.2 Future Work

There are several issues that can be explored further from this thesis. These are:

1) Possibility to use this approach to compare human generated summary and automatic generated summary. To compare among summaries, we need to extract the concepts from both human generated summary and automatic generated summary. Then, apply the semantic measures on the extracted concepts.

2) Using other semantic measures such as path measure, Jiang & Conrath measure and context Vectors measure.

3) Apply semantic measures for Arabic text. It's more difficult than English language because we must have an Arabic version of WordNet. Also, we need semantic measures that are designed to compute the similarity for Arabic concepts.

4) Use semantic measures to check relevancy and plagiarism. The plagiarism is already exist but with exact matching. Our idea is to apply the semantic matching using the semantic measures to check the relevancy and plagiarism.

References

- Alguliev, R., & Aliguliyev, R. (2007). Experimental investigating the F-measure as similarity measure for automatic text summarization. *Applied and Computational Mathematics*, *Vol.6* (No.2), PP. 278-287.

- Al-Khiaty, M., & Ahmed, M. (2015). UML Class Diagrams: Similarity Aspects and Matching. *LNSE*, *4*(1), 41-47. http://dx.doi.org/10.7763/lnse.2015.v4.221

- Baader, F., Sertkaya, B., & Turhan, A. (2007). Computing the least common subsumer w.r.t. a background terminology. *Journal of Applied Logic*, *5*(3), 392-420. http://dx.doi.org/10.1016/j.jal.2006.03.002

- Barzilay, R., & Elhadad, M. (1999). Using lexical chains for text summarization. *Advances in Automatic Text Summarization*, PP. 111-121.

- Boon young, T., & Mingkhwan, A. (2015). Document Similarity using Computer Science Ontology based on Edge Counting and N-Grams, pp. 23-24.

- Buckley, J., Liu, G., Wang, R., & Zhou, H. (2011). A WordNet-based Semantic Similarity Measure Enhanced by Internet-based Knowledge. *In SEKE,* pp. 175-178.

- Chai, T., & Draxler, R. (2014). Root mean square error (RMSE) or mean absolute error (MAE) Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247-1250. http://dx.doi.org/10.5194/gmd-7-1247-2014.

- Corley, C., & Mihalcea, R. (2005). Measuring the semantic similarity of texts. *In Proceedings of The ACL Workshop On Empirical Modeling Of Semantic Equivalence And Entailment. Association for Computational Linguistics.*pp. 13-18.

- Fukusima, T., & Okumura, M. (2001). Text Summarization Challenge Text

Summarization Evaluation in Japan. *In Proceedings of The Workshop On Automatic Summarization. Pittsburgh, PA, USA*.

- Gupta, V. (2014). A Survey of Various Summary Evaluation Techniques. *International Journal of Advanced Research In Computer Science And Software Engineering*, *Vol. 4*(ISSN: 2320-088X).
- HOUSE, D., KLEIN, G., MANI, I., HIRSCHMAN, L., FIRMIN, T., & SUNDHEIM, B. (2002). SUMMAC: a text summarization evaluation. *Natural Language Engineering*, *8*(01). http://dx.doi.org/10.1017/s1351324901002741
- Jiang, R., Dou, X., & Gan, M. (2013). From Ontology to Semantic Similarity: Calculation of Ontology-Based Semantic Similarity. *The Scientific World Journal*, *2013*, 1-11. http://dx.doi.org/10.1155/2013/793091
- Kayed, A., Hirzallah, N., & Alharibat, M. (2013). Ontological Concepts Coverage Measure for Software- Component Description. *European Journal of Scientific Research*, PP.77-99.
- Kayed, A., Nizar, M., & Alfayoumi, M. (2010). Ontology concepts for requirements engineering process in e-government applications. *In Internet and Web Applications and Services (ICIW). Fifth International Conference on IEEE*, pp. 396-400.
- Lin, C. (2004). Looking for a few good metrics: ROUGE and its evaluation. *NTCIR Workshop*.
- Liu, G., Wang, R., Buckley, J., & Zhou, H. (2011). A WordNet-based Semantic Similarity Measure Enhanced by Internet-based Knowledge. *In SEKE*, pp. 175-178.
- Maedche, A., & Volz, R. (2001). The ontology extraction & maintenance framework Text-To-Onto. *In Proc. Workshop on Integrating Data Mining and Knowledge* Management, USA, 1-12.

- Mani, I. (2001a). Recent developments in text summarization. *Inproceedings of The Tenth International Conference On Information And Knowledge Management.ACM*, PP. 529-531.

- Mani, I. (2001b). Summarization Evaluation: An overview. *In proceedings of The Tenth International Conference on Information and Knowledge Management*.

- McInnes, B., & Pedersen, T. (2013). Evaluating measures of semantic similarity and relatedness to disambiguate terms in biomedical text. *Journal of Biomedical Informatics*, *46*(6), 1116-1124. http://dx.doi.org/10.1016/j.jbi.2013.08.008

- Meng, L., Huang, R., & Gu, J. (2013). A review of semantic similarity measures in wordnet. *International Journal of Hybrid Information Technology*, *Vol 6(1)*, PP. 1-12.

- Michelizzi, J. (2005). Semantic relatedness applied to all words sense disambiguation, *(Doctoral dissertation)*, *University of Minnesota*, *Twin Cities.*

- Mittal, K., & Jain, A. (2015). WORD SENSE DISAMBIGUATION METHOD USING SEMANTIC SIMILARITY MEASURES AND OWA OPERATOR, *Vol.5*, *No.2*.

- Moghaddas, B., Kahani, M., Toosi, S., Pourmasoumi, A., & Estiri, A. (2013). Pasokh: A standard corpus for the evaluation of Persian text summarizers. *In Computer and Knowledge Engineering (ICCKE), 2013 3Th International Econference On IEEE*, pp. 471-475.

- Morris, A., Kasper, G., & Adams, D. (1992). The Effects and Limitations of Automated Text Condensing on Reading Comprehension Performance. *Information Systems Research*, *3*(1), pp.17-35.

- Pedersen, T., Patwardhan, S., & Michelizzi, J. (2004). WordNet:: Similarity: measuring the relatedness of concepts. *Association for Computational Linguistics*, pp. 38-41.

- Seddiqui, M., & Aono, M. (2010). Metric of intrinsic information content for measuring semantic similarity in ontology. *In Proceedings Of The Seventh Asia-Pacific Conference On Conceptual Modelling, Vol. 110*, pp. 89-96.

- Slimani, T. (2013). Description and Evaluation of Semantic Similarity Measures Approaches. *International Journal of Computer Applications*, *80*(10), 25-33. http://dx.doi.org/10.5120/13897-1851

- Steinberger, J., & Jezek, K. (2009). Evaluation measures for text summarization. *Computing and Informatics*, *Vol.28* (No.2), 1001-1026.

Appendix

Appendix

1. NIST assigned agreement

Organization Application
to use the

TIPSTER Information-Retrieval Text Research Collections

The Computer Science Dept., an organization of approximately _________ people engaging in research and development of natural-language-processing, information-retrieval or document-understanding systems, which is a part of:

Corporation/Partnership/Legal Entity IT Faculty , MEU
Official mail address AMMan - Jordan
Queen Alia AirPort street , P.O. Box:- 11831 AMMan

Telephone 00 962 6 479 0222
Facsimile 00 962 6 412 9613
Email Sec-IT @ meu.edu.jo

apply(ies) to use the information designated as the TIPSTER Information-Retrieval Text Research Collection subject to the following understandings, terms and conditions. These understandings, terms and conditions apply equally to all or to part of the information.

Permitted Uses

1. The information may only be used for research and development of natural-language-processing, information-retrieval or document-understanding systems.
2. Summaries, analyses and interpretations of the linguistic properties of the information may be derived and published, provided it is not possible to reconstruct the information from these summaries.
3. Small excerpts of the information may be displayed to others or published in a scientific or technical context, solely for the purpose of describing the research and development and related issues. Any such use shall not infringe on the rights of any third party including, but limited to, the authors and publishers of the excerpts.

Access to the Information by Individuals

1. Access to the information by an individual person is to be controlled by that person's organization. The organization may only grant access to people working under its control, i.e., its own members, consultants to the organization, or individuals providing service to the organization.
2. Individuals may be allowed access to the information only after completion and submission of the Individual Application form. The access is to be terminated when the conditions of the application no longer apply. The organization will retain the applications of all persons ever granted access to the information and make them

available upon request to any of the copyright holders and to the institution or agency holding this completed application.

3. The organization will maintain and post a list of people with current and recently-terminated access to the information.
4. An individual with access may only display the information to or share the information with persons whom his organization lists as having access to the information.

Copyright

1. The copyright holders retain ownership and reserve all rights pertaining to the use and distribution of the information.
2. Except as specifically permitted above and as necessary to use and maintain the integrity of the information on computers used by the organization; the display, reproduction, transmission, distribution or publication of the information is prohibited. Violations of the copyright restrictions on the information may result in legal liability.
3. Copyright holders of the information contained in the Information-Retrieval Text Research Collection include:
   - Wall Street Journal material, copyright 1987-1992 Dow Jones, Inc.
   - AP material, copyright 1988-1990 Associated Press
   - San Jose Mercury News material, copyright 1991 San Jose Mercury News
   - Computer Library materials copyright owned by Ziff Communications Company
   - and other parties as their respective interests may appear.

By the Organization:

Signature ____________________

Date 29-4-2015

Name *(please print)* prof. Ahmad Kayed

Title IT Dean

Accepted by NIST:

Signature ____________________

Date ____________________

Name *(please print)* ____________________

Title ____________________

Please complete and email a PDF file of this form to:

lori.buckland@nist.gov

Organization Application
to use the

TREC Information-Retrieval Text Research Collections

The Computer Science Dept., an organization of approximately ____________ people engaging in research and development of natural-language-processing, information-retrieval or document-understanding systems, which is a part of:

Corporation/Partnership/Legal Entity IT Faculty, MEU
Official mail address Amman - Jordan
Queen Alia AirPort street, P.O. Box: - 11831 AMMan

Telephone 00 962 6 479 0222
Facsimile 00 962 6 412 9613
Email Sec-IT@meu.edu.jo

apply(ies) to use the information designated as the TREC Information-Retrieval Text Research Collection subject to the following understandings, terms and conditions. These understandings, terms and conditions apply equally to all or to part of the information.

Permitted Uses

1. The information may only be used for research and development of natural-language-processing, information-retrieval or document-understanding systems.
2. Summaries, analyses and interpretations of the linguistic properties of the information may be derived and published, provided it is not possible to reconstruct the information from these summaries.
3. Small excerpts of the information may be displayed to others or published in a scientific or technical context, solely for the purpose of describing the research and development and related issues. Any such use shall not infringe on the rights of any third party including, but limited to, the authors and publishers of the excerpts.

Access to the Information by Individuals

1. Access to the information by an individual person is to be controlled by that person's organization. The organization may only grant access to people working under its control, i.e., its own members, consultants to the organization, or individuals providing service to the organization.
2. Individuals may be allowed access to the information only after completion and submission of the Individual Application form. The access is to be terminated when the conditions of the application no longer apply. The organization will retain the applications of all persons ever granted access to the information and make them available upon request to any of the copyright holders and to the institution or agency holding this completed application.

3. The organization will maintain and post a list of people with current and recently-terminated access to the information.
4. An individual with access may only display the information to or share the information with persons whom his organization lists as having access to the information.

Copyright

1. The copyright holders retain ownership and reserve all rights pertaining to the use and distribution of the information.
2. Except as specifically permitted above and as necessary to use and maintain the integrity of the information on computers used by the organization; the display, reproduction, transmission, distribution or publication of the information is prohibited. Violations of the copyright restrictions on the information may result in legal liability.
3. Copyright holders of the information contained in the Information-Retrieval Text Research Collection include:
   - Financial Times material, copyright 1988-1990, 1992-1994 Financial Times Limited;
   - Foreign Broadcast Information Service material, copyright 1994;
   - Los Angeles Times material, copyright Los Angeles Times 1989-1990;
   - Content provided by Wisers Information Limited, includes:
     - Takongpao material, copyright 1998-1999
     - Hong Kong Commercial Daily, copyright 1998-1999
     - Hong Kong Daily News, copyright 1998-1999;
   - and other parties as their respective interests appear.

By the Organization:

Signature ____________________

Date 29-4-2015

Name (please print) Prof. Ahmad K. A. Kayed

Title prof. IT Dean

Accepted by NIST:

Signature ____________________

Date ____________________

Name (please print) ____________________

Title ____________________

Please complete and email a PDF file of this form to:

lori.buckland@nist.gov

In your email message include:

2. Table of original extracted concepts in file.

No.	Concepts	No.	Concepts	No.	Concepts	No.	Concepts	No.	Concepts
1	Defense	**28**	Prosecution	**55**	Order	**82**	Guard	**109**	Television
2	Minister	**29**	Institution	**56**	Advice	**83**	Stuff	**110**	Interview
3	Priest	**30**	Power	**57**	Government	**84**	Idea	**111**	Division
4	Magistrate	**31**	Abuse	**58**	Lawmaker	**85**	Business	**112**	Riot
5	Murder	**32**	Constitution	**59**	Bed	**86**	Army	**113**	Police
6	Day	**33**	Statement	**60**	Result	**87**	People	**114**	Police
7	Embassy	**34**	Word	**61**	Water	**88**	Administration	**115**	Line
8	Police	**35**	Military	**62**	Cannon	**89**	Government	**116**	spokesman
9	Source	**36**	Support	**63**	Decade	**90**	Report	**117**	Response
10	Insult	**37**	Portrait	**64**	Application	**91**	Disappearance	**118**	Death
11	Team	**38**	Extradition	**65**	Regime	**92**	Passport	**119**	Pressure
12	Claim	**39**	Law	**66**	Delegation	**93**	Family	**120**	Release
13	Arrest	**40**	Protester	**67**	Place	**94**	Total	**121**	Injury
14	Warrant	**41**	Arrest	**68**	Politician	**95**	Police	**122**	Violation
15	Attorney	**42**	Opponent	**69**	Scene	**96**	Gas	**123**	Attitude
16	Envoy	**43**	Rightist	**70**	Decision	**97**	Allegation	**124**	Lobby
17	Extradition	**44**	Politician	**71**	Judge	**98**	News	**125**	Group
18	Request	**45**	Time	**72**	Evening	**99**	Court	**126**	Placard
19	Opinion	**46**	Amnesty	**73**	Dictator	**100**	Strongman	**127**	Bearing
20	Term	**47**	Month	**74**	Iron	**101**	Rights	**128**	Condition
21	Rally	**48**	Father	**75**	Fist	**102**	Abuse	**129**	Son
22	Week	**49**	Immunity	**76**	Visitor	**103**	Country	**130**	Year
23	Boss	**50**	Caption	**77**	Diplomat	**104**	Demonstrator	**131**	demonstration
24	Government	**51**	Capital	**78**	Policy	**105**	Operation	**132**	anonymity
25	Official	**52**	Publicity	**79**	Trial	**106**	Official	**133**	Reign
26	Post	**53**	Patient	**80**	Black	**107**	Event	**134**	Hospital
27	crime	**54**	Arrest	**81**	Police	**108**	Dissident	**135**	Protest

3. semantic matching results for T1

A) WuP Measure

No	OC	CC1	CC2	CC3	CC4	CC5	CC6	CC7	CC8	CC9	CC10	CC11	CC12	CC13	CC14	CC15	CC16	CC17	CC18	CC19	CC20
		dictator	confusion	hospital	arrest	murder	citizen	rule	legality	move	supporter	immunity	extradition	outcome	request	charge	community	surgery	death	court	government
1	defense	0.5556	0.6667	0.7778	0.6667	0.5455	0.5556	0.7778	0.4615	0.7778	0.7059	0.5455	0.6364	0.625	0.7778	0.8889	0.75	0.7778	0.625	0.8235	0.8
2	minister	0.7273	0.6	0.4545	0.6	0.5	0.7273	0.7	0.3333	0.7	0.7826	0.5	0.5	0.5263	0.6316	0.7	0.4	0.6667	0.5263	0.6667	0.6
3	priest	0.7619	0.6667	0.4762	0.6667	0.2857	0.7619	0.625	0.5333	0.3529	0.7619	0.6667	0.3	0.4286	0.4286	0.6957	0.5882	0.4762	0.7143	0.6957	0.375
4	magistrate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	murder	0.2	0.6667	0.2727	0.5455	1	0.2	0.5455	0.3	0.7273	0.2	0.7692	0.4615	0.4762	0.5714	0.5455	0.3333	0.5217	0.88	0.6087	0.5455
6	day	0.7826	0.625	0.4762	0.625	0.3158	0.7619	0.8	0.5333	0.4	0.7619	0.625	0.3333	0.4286	0.5	0.6957	0.5556	0.4762	0.6667	0.6957	0.4286
7	embassy	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	police	0.2667	0.4	0.7059	0.4	0.2857	0.2667	0.4	0.4	0.3529	0.2667	0.4	0.3	0.375	0.4286	0.4	0.7059	0.3333	0.4286	0.7059	0.75
9	source	0.8182	0.4444	0.6667	0.4444	0.3636	0.8	0.7778	0.4	0.4444	0.8	0.4	0.381	0.6667	0.5714	0.7273	0.6667	0.6667	0.5714	0.7273	0.6
10	insult	0.2857	0.6316	0.375	0.6316	0.5217	0.2857	0.7368	0.4286	0.7368	0.2857	0.5217	0.5217	0.5556	0.7692	0.7368	0.5	0.7	0.5556	0.6	0.6316
11	team	0.2857	0.4286	0.7059	0.4286	0.3	0.2857	0.4286	0.4286	0.375	0.2857	0.4286	0.3158	0.4	0.4615	0.4286	0.7059	0.3529	0.4615	0.7059	0.75
12	claim	0.2667	0.6667	0.3529	0.6667	0.5455	0.2667	0.8	0.4	0.7778	0.2667	0.5455	0.5455	0.5882	0.8571	0.8889	0.4615	0.7368	0.5882	0.6316	0.6667
13	arrest	0.2857	0.7143	0.375	1	0.5455	0.2857	0.6667	0.5714	0.6667	0.2857	0.7143	0.5455	0.5882	0.7059	0.6667	0.625	0.6316	0.7692	0.6316	0.6667
14	warrant	0.2857	0.4286	0.375	0.4286	0.3	0.2857	0.7143	0.4286	0.375	0.2857	0.4286	0.3158	0.4	0.7692	0.7059	0.5714	0.3529	0.4615	0.4	0.4
15	attorney	0.7619	0.2667	0.4762	0.2667	0.1905	0.7619	0.4348	0.2667	0.2353	0.7619	0.2667	0.2	0.4286	0.2857	0.6957	0.4211	0.4762	0.375	0.6957	0.25
16	envoy	0.75	0.3529	0.4545	0.3529	0.2609	0.7273	0.4706	0.3529	0.3158	0.7826	0.3529	0.2727	0.4	0.5	0.6667	0.4	0.4545	0.375	0.6667	0.3333
17	extradition	0.2105	0.5455	0.2857	0.5455	0.4615	0.2105	0.5455	0.3158	0.5455	0.2105	0.4615	1	0.5	0.5714	0.5455	0.3529	0.5217	0.5	0.5217	0.6364
18	request	0.3077	0.7059	0.4	0.7059	0.5714	0.3077	0.7692	0.4615	0.7059	0.3077	0.5714	0.5714	0.625	1	0.9412	0.5455	0.6667	0.625	0.6667	0.7059
19	opinion	0.3077	0.6316	0.4	0.6316	0.5217	0.3077	0.8235	0.4615	0.6316	0.3077	0.5217	0.6087	0.5882	0.8333	0.7143	0.5455	0.6	0.5882	0.6	0.7368
20	term	0.4545	0.4706	0.5455	0.4706	0.3478	0.4545	0.8421	0.4706	0.4211	0.5714	0.4706	0.3636	0.4444	0.7692	0.8	0.5714	0.5455	0.7692	0.5714	0.5714
21	rally	0.375	0.7059	0.625	0.7059	0.6087	0.375	0.7059	0.4286	0.7368	0.375	0.6087	0.6667	0.6667	0.75	0.7059	0.8571	0.6667	0.6667	0.8	0.8235
22	week	0.2857	0.4286	0.375	0.4286	0.3	0.2857	0.8	0.4286	0.375	0.2857	0.4286	0.3158	0.4	0.4615	0.4286	0.5	0.3529	0.5714	0.4	0.4
23	boss	0.8	0.2857	0.7368	0.2857	0.2	0.8	0.5714	0.2857	0.25	0.8	0.2857	0.2105	0.4615	0.3077	0.7273	0.4706	0.7368	0.4	0.7778	0.2667
24	government	0.2667	0.6667	0.7059	0.6667	0.5455	0.2667	0.6667	0.4	0.6667	0.2667	0.5455	0.6364	0.625	0.7059	0.6667	0.7059	0.75	0.625	0.9412	1
25	official	0.8	0.2857	0.5	0.2857	0.2	0.8	0.4545	0.2857	0.25	0.8571	0.2857	0.2105	0.4615	0.3077	0.7273	0.4444	0.5	0.4	0.7273	0.2667
26	post	0.7826	0.7059	0.6316	0.7059	0.64	0.7619	0.7778	0.4615	0.7778	0.8182	0.64	0.5714	0.625	0.75	0.7778	0.7273	0.7368	0.6667	0.6957	0.7059
27	crime	0.2353	0.6316	0.3158	0.6316	0.5217	0.2353	0.7368	0.3529	0.7368	0.2353	0.5217	0.5217	0.5556	0.6667	0.7368	0.4	0.7	0.5556	0.6	0.6316
28	prosecution	0.3077	0.6667	0.5333	0.6667	0.5455	0.3077	0.7778	0.4615	0.7778	0.3077	0.5455	0.5833	0.5882	0.7059	0.7778	0.7273	0.7368	0.5882	0.6316	0.7
29	institution	0.5263	0.7273	0.9524	0.6	0.75	0.5263	0.8	0.4286	0.8	0.6667	0.75	0.5	0.5263	0.6316	0.6	0.75	0.7368	0.7826	0.7778	0.8
30	power	0.7273	0.7692	0.6316	0.7692	0.4211	0.7273	0.7692	0.7692	0.5333	0.7273	0.8	0.4444	0.7143	0.5714	0.8235	0.6667	0.5263	0.8333	0.6667	0.75

63	decade	0.2857	0.4286	0.375	0.4286	0.3	0.2857	0.8	0.4286	0.375	0.2857	0.4286	0.3158	0.4	0.4615	0.4286	0.5	0.3529	0.5714	0.4	0.4	.6316	0.6667
64	application	0.5	0.7368	0.375	0.7059	0.6667	0.5	0.7778	0.4286	0.8235	0.5	0.6667	0.5714	0.625	0.9231	0.7778	0.5	0.7368	0.7	0.7778	0.7059	.6667	0.6
65	regime	0.2667	0.4444	0.7059	0.4444	0.3636	0.2667	0.7778	0.4	0.4444	0.2667	0.4	0.381	0.4706	0.4706	0.4706	0.7059	0.5455	0.5263	0.9412	1	.6667	0.7059
66	delegation	0.2857	0.6	0.75	0.6	0.5	0.2857	0.6	0.4286	0.6	0.2857	0.5	0.5833	0.5556	0.6316	0.6	0.75	0.5714	0.5556	0.75	0.8	.6316	0.6667
67	place	0.4706	0.75	0.4706	0.7143	0.5455	0.4706	0.8	0.5714	0.7778	0.5	0.8571	0.5455	0.5882	0.7059	0.8182	0.75	0.7368	0.7692	0.8	0.6667	-	-
68	politician	0.8	0.2857	0.5	0.2857	0.2	0.8	0.4545	0.2857	0.25	0.8	0.2857	0.2105	0.4615	0.3077	0.7273	0.4444	0.5	0.4	0.7273	0.2667	.7778	0.7059
69	scene	0.5	0.7143	0.6	0.7143	0.4762	0.5	0.6667	0.5714	0.5882	0.6316	0.7143	0.5	0.75	0.625	0.7	0.75	0.6	0.7692	0.7368	0.625	0.6	0.3529
70	decision	0.2857	0.7	0.375	0.6667	0.6364	0.2857	0.6667	0.5714	0.9474	0.2857	0.6364	0.5455	0.9412	0.7059	0.6667	0.5	0.6316	0.7059	0.7368	0.6667	.5217	0.6364
71	judge	0.8	0.2857	0.5	0.2857	0.2	0.8	0.4545	0.2857	0.25	0.8182	0.2857	0.2105	0.4615	0.3077	0.7273	0.4444	0.5	0.4	0.7273	0.2667	.7059	0.75
72	evening	0.2857	0.4286	0.375	0.4286	0.3	0.2857	0.8	0.4286	0.375	0.2857	0.4286	0.3158	0.4	0.4615	0.4286	0.5	0.3529	0.5714	0.4	0.4	.7619	0.2857
73	dictator	1	0.2857	0.5	0.2857	0.2	0.8	0.4545	0.2857	0.25	0.8	0.2857	0.2105	0.4615	0.3077	0.7273	0.4444	0.5	0.4	0.7273	0.2667	.6316	0.6667
74	iron	0.5263	0.375	0.6316	0.375	0.2727	0.5263	0.6667	0.375	0.3333	0.6957	0.375	0.2857	0.4286	0.4	0.6316	0.5333	0.6316	0.4	0.6667	0.3529	.8182	0.2857
75	fist	0.3529	0.2353	0.3	0.2353	0.1739	0.3529	0.2727	0.2353	0.2105	0.3529	0.2353	0.1818	0.375	0.25	0.3158	0.3333	0.3	0.3333	0.3158	0.2222	.6957	0.25
76	visitor	0.8	0.2857	0.5	0.2857	0.2	0.8	0.4545	0.2857	0.25	0.8	0.2857	0.2105	0.4615	0.3077	0.7273	0.4444	0.5	0.4	0.7273	0.2667	.7273	0.2667
77	diplomat	0.75	0.25	0.4545	0.25	0.1818	0.7273	0.4167	0.25	0.2222	0.8182	0.25	0.1905	0.4	0.2667	0.6667	0.4	0.4545	0.3529	0.6667	0.2353	.5556	0.625
78	policy	0.2353	0.4211	0.3158	0.4211	0.3478	0.2353	0.7368	0.3529	0.4211	0.2353	0.3529	0.3636	0.4444	0.625	0.6667	0.4	0.5217	0.5	0.4	0.5714	-	-
79	trial	0.2667	0.6667	0.3529	0.6667	0.5455	0.2667	0.7778	0.4	0.7778	0.2667	0.5455	0.6087	0.7059	0.7059	0.7778	0.4615	0.7368	0.7059	0.6316	0.7368	.4286	0.4286
80	black	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	.7273	0.5217
81	police	0.2667	0.4	0.7059	0.4	0.2857	0.2667	0.4	0.4	0.3529	0.2667	0.4	0.3	0.375	0.4286	0.4	0.7059	0.3333	0.4286	0.7059	0.75	.6087	0.5455
82	guard	0.7619	0.6316	0.6667	0.6316	0.5385	0.7619	0.7619	0.5333	0.8696	0.7619	0.625	0.5217	0.5556	0.6667	0.9091	0.6667	0.7	0.5714	0.8333	0.7059	0.6	0.6316
83	stuff	0.5333	0.6154	0.5333	0.6154	0.381	0.5333	0.7143	0.7692	0.4706	0.5714	0.6154	0.4	0.5455	0.7692	0.8235	0.6154	0.5714	0.6667	0.5714	0.6316	.8421	0.8889
84	idea	0.2857	0.5	0.375	0.5	0.4	0.2857	0.875	0.4286	0.5	0.2857	0.4286	0.4211	0.5333	0.5714	0.5333	0.5	0.6	0.5882	0.4706	0.6667	.4286	0.4286
85	business	0.3077	0.7059	0.7059	0.7059	0.5714	0.3077	0.8235	0.4615	0.8235	0.3077	0.5714	0.6957	0.625	0.75	0.8235	0.7273	0.7778	0.625	0.7059	0.75	.7273	0.4706
86	army	0.2667	0.4	0.6667	0.4	0.2857	0.2667	0.4	0.4	0.3529	0.2667	0.4	0.3	0.375	0.4286	0.4	0.8	0.3333	0.4286	0.75	0.7059	.6316	0.6667
87	people	0.3333	0.5	0.5714	0.5	0.3333	0.3333	0.5	0.5	0.4286	0.3333	0.5	0.3529	0.4615	0.5455	0.5	0.8	0.4	0.5455	0.6154	0.6154	.7059	0.75
88	administration	0.2857	0.6667	0.625	0.6667	0.5455	0.2857	0.7059	0.4286	0.6667	0.2857	0.5455	0.6364	0.625	0.7059	0.6667	0.7143	0.6316	0.625	0.6667	1	0.4	0.4
89	government	0.2667	0.6667	0.7059	0.6667	0.5455	0.2667	0.6667	0.4	0.6667	0.2667	0.5455	0.6364	0.625	0.7059	0.6667	0.7059	0.75	0.625	0.9412	1	.9412	1
90	report	0.2857	0.6667	0.375	0.6667	0.5455	0.2857	0.7143	0.4286	0.6667	0.2857	0.5455	0.5455	0.7059	0.8235	0.7778	0.5	0.6316	0.7059	0.6316	0.6667	.7273	0.2667
91	disappearance	0.2857	0.7273	0.375	0.7059	0.8333	0.2857	0.7059	0.4286	0.8	0.2857	0.8333	0.5714	0.8	0.75	0.7059	0.5	0.6667	0.8696	0.6667	0.7059	0.8	0.2667
92	passport	0.2857	0.5714	0.375	0.5714	0.3	0.2857	0.6667	0.7143	0.375	0.2857	0.5714	0.3158	0.4	0.7143	0.625	0.5	0.3529	0.6154	0.4	0.4	.5556	0.625
93	family	0.8	0.4615	0.7059	0.4615	0.3158	0.8	0.4615	0.4615	0.4	0.8	0.4615	0.3333	0.4615	0.5	0.7273	0.7273	0.5	0.5	0.7273	0.75	.7059	0.375
62	cannon	0.4762	0.6957	0.5714	0.5714	0.64	0.4762	0.75	0.3158	0.8571	0.6364	0.64	0.48	0.5	0.6	0.5714	0.4211	0.5714	0.6667	0.6364	0.5714	0.6364	0.5714

94	total	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
95	police	0.2667	0.4	0.7059	0.4	0.2857	0.2667	0.4	0.4	0.3529	0.2667	0.4	0.3	0.375	0.4286	0.4	0.7059	0.3333	0.4286	0.7059	0.75
96	gas	0.4615	0.7059	0.5455	0.6667	0.2857	0.4615	0.75	0.5333	0.3529	0.5714	0.8	0.3	0.6667	0.4286	0.7273	0.5882	0.5455	0.7143	0.5714	0.375
97	allegation	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
98	news	0.2857	0.5556	0.375	0.5556	0.4545	0.2857	0.7143	0.6667	0.5556	0.2857	0.5333	0.4762	0.5882	0.7692	0.6667	0.5	0.5263	0.5882	0.5263	0.5882
99	court	0.7273	0.6667	0.8	0.6316	0.6087	0.7273	0.6316	0.4	0.7368	0.7273	0.6087	0.5217	0.5556	0.6667	0.6667	0.8	0.9	0.6364	1	0.9412
100	strongman	0.9524	0.2667	0.4762	0.2667	0.1905	0.7619	0.4348	0.2667	0.2353	0.7619	0.2667	0.2	0.4286	0.2857	0.6957	0.4211	0.4762	0.375	0.6957	0.25
101	rights	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
102	abuse	0.2857	0.6667	0.375	0.6667	0.5455	0.2857	0.7778	0.4286	0.7778	0.2857	0.5455	0.5455	0.5882	0.7692	0.7778	0.5	0.7368	0.5882	0.6316	0.6667
103	country	0.4706	0.4615	0.6667	0.4615	0.3158	0.4706	0.4615	0.4615	0.4	0.5	0.4615	0.3333	0.4615	0.5	0.4615	0.8235	0.4706	0.5	0.7778	0.7059
104	demonstrator	0.7619	0.2667	0.4762	0.2667	0.1905	0.7619	0.4348	0.2667	0.2353	0.8182	0.2667	0.2	0.4286	0.2857	0.6957	0.4211	0.4762	0.375	0.6957	0.25
105	operation	0.5	0.7143	0.4	0.7143	0.5714	0.5	0.8421	0.5714	0.8235	0.5	0.7143	0.5714	0.7273	0.75	0.8889	0.625	1	0.7692	0.6667	0.7059
106	official	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
107	event	0.4615	0.7143	0.4286	0.7143	0.5556	0.4615	0.75	0.5333	0.7143	0.4615	0.8	0.5882	1	0.7692	0.8235	0.6	0.6667	0.7692	0.6667	0.7692
108	dissident	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
109	television	0.4762	0.2222	0.5714	0.2222	0.1667	0.4762	0.6087	0.2222	0.2	0.6	0.2222	0.1739	0.3529	0.2353	0.5455	0.4211	0.5714	0.3158	0.6	0.2105
110	interview	0.25	0.6	0.3333	0.6	0.5	0.25	0.6	0.375	0.6	0.25	0.5	0.5	0.5263	0.8421	0.8	0.4286	0.5714	0.5263	0.5714	0.6
111	division	0.3077	0.8182	0.6667	0.6316	0.6667	0.3077	0.8889	0.4615	0.8	0.3077	0.6667	0.5217	0.5556	0.7778	0.7368	0.7273	0.7273	0.6957	0.6667	0.7059
112	riot	0.2857	0.8571	0.375	0.7143	0.6087	0.2857	0.6667	0.5714	0.7368	0.2857	0.7143	0.5217	0.5556	0.7143	0.6667	0.625	0.6364	0.7692	0.7	0.6316
113	police	0.2667	0.4	0.7059	0.4	0.2857	0.2667	0.4	0.4	0.3529	0.2667	0.4	0.3	0.375	0.4286	0.4	0.7059	0.3333	0.4286	0.7059	0.75
114	police	0.2667	0.4	0.7059	0.4	0.2857	0.2667	0.4	0.4	0.3529	0.2667	0.4	0.3	0.375	0.4286	0.4	0.7059	0.3333	0.4286	0.7059	0.75
115	line	0.5882	0.7059	0.7059	0.7059	0.5714	0.5882	0.8235	0.6154	0.8235	0.75	0.6154	0.6087	0.625	0.75	0.8235	0.7143	0.7778	0.6667	0.75	0.7368
116	spokesman	0.7619	0.2667	0.4762	0.2667	0.1905	0.7619	0.4348	0.2667	0.2353	0.8571	0.2667	0.2	0.4286	0.2857	0.6957	0.4211	0.4762	0.375	0.6957	0.25
117	response	0.4286	0.7059	0.375	0.7059	0.5714	0.4286	0.7143	0.5714	0.7059	0.4286	0.6667	0.5714	0.9231	0.875	0.8235	0.5	0.6667	0.6667	0.6667	0.7059
118	death	0.4	0.7692	0.4	0.7692	0.88	0.4	0.7143	0.6154	0.7619	0.4	0.8	0.5	0.75	0.625	0.7059	0.6667	0.5556	1	0.6364	0.625
119	pressure	0.4	0.6667	0.3333	0.6667	0.5455	0.4	0.6667	0.625	0.6667	0.4	0.7059	0.5455	0.7143	0.7059	0.8235	0.5263	0.6316	0.75	0.6316	0.6667
120	release	0.5263	0.7273	0.6316	0.7059	0.8333	0.5263	0.8235	0.4	0.9	0.7	0.96	0.5714	0.7059	0.75	0.8235	0.4706	0.7778	0.9412	0.7368	0.7059
121	injury	0.25	0.6667	0.3333	0.6667	0.5455	0.25	0.7778	0.4706	0.7778	0.25	0.7059	0.5455	0.6316	0.7059	0.7778	0.5263	0.7368	0.7368	0.6316	0.6667
122	violation	0.25	0.6667	0.3333	0.6667	0.5455	0.25	0.7778	0.375	0.7778	0.25	0.5455	0.5455	0.5882	0.7059	0.7778	0.4286	0.7368	0.5882	0.6316	0.6667
123	attitude	0.3077	0.5714	0.4	0.5714	0.4615	0.3077	0.7692	0.5714	0.6364	0.3077	0.6667	0.4615	0.5714	0.5714	0.6364	0.5714	0.6087	0.625	0.5217	0.625
124	lobby	0.5	0.4615	0.7	0.4615	0.3158	0.5	0.5455	0.4615	0.4	0.6316	0.4615	0.3333	0.4286	0.5	0.5714	0.7273	0.9	0.5	0.9	0.7059
125	group	0.4615	0.5455	0.6154	0.5455	0.3529	0.4615	0.5455	0.5455	0.4615	0.4615	0.5455	0.375	0.5	0.6	0.5455	0.8889	0.4286	0.6	0.6667	0.6667

126	placard	0.3077	0.4615	0.4	0.4615	0.3158	0.3077	0.6154	0.4615	0.4	0.3077	0.4615	0.3333	0.4286	0.6667	0.5714	0.5455	0.375	0.5	0.4286	0.4286
127	bearing	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
128	condition	0.3077	0.8	0.4	0.7692	0.5455	0.3077	0.8889	0.6154	0.7778	0.3077	0.9231	0.5455	0.5882	0.7692	0.8	0.6667	0.8421	0.8333	0.6316	0.6667
129	son	0.7619	0.381	0.4762	0.381	0.32	0.7619	0.5714	0.3158	0.381	0.7619	0.32	0.3333	0.4286	0.4	0.6957	0.4211	0.48	0.4545	0.6957	0.5217
130	year	0.2857	0.4286	0.625	0.4286	0.3	0.2857	0.8	0.4286	0.375	0.2857	0.4286	0.3158	0.4	0.4615	0.4286	0.8571	0.3529	0.5714	0.8	0.6667
131	demonstration	0.3077	0.6316	0.4	0.6316	0.5217	0.3077	0.7	0.4615	0.7	0.3077	0.5217	0.6087	0.5882	0.6667	0.7	0.5455	0.6667	0.5882	0.6	0.7368
132	anonymity	0.25	0.625	0.3333	0.625	0.2727	0.25	0.5882	0.5	0.3333	0.25	0.625	0.2857	0.3529	0.4	0.5263	0.5556	0.3158	0.6667	0.3529	0.3529
133	reign	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
134	hospital	0.5	0.375	1	0.375	0.2727	0.5	0.5455	0.375	0.3333	0.6316	0.375	0.2857	0.375	0.4	0.5714	0.6667	0.7	0.4	0.8	0.7059
135	protest	0.2667	0.6667	0.3529	0.6667	0.5455	0.2667	0.6667	0.4	0.6667	0.2667	0.5455	0.6364	0.625	0.8235	0.7778	0.4615	0.6316	0.625	0.6316	0.7778

B) LCH measure

No	OC	CC1	CC2	CC3	CC4	CC5	CC6	CC7	CC8	CC9	CC10	CC11	CC12	CC13	CC14	CC15	CC16	CC17	CC18	CC19	CC20
		dictator	confusion	hospital	arrest	murder	citizen	rule	legality	move	supporter	immunity	extradition	outcome	request	charge	community	surgery	death	court	government
1	defense	1.4917	1.743	2.0794	1.743	1.291	1.4917	2.079	1.6094	2.079	1.8971	1.6094	1.4917	1.8971	2.0794	2.59	2.3026	2.0794	1.743	2.303	2.3026
2	minister	1.743	1.4917	1.1239	1.4917	1.1239	1.743	1.743	1.1239	1.743	1.8971	1.1239	1.1239	1.3863	1.6094	1.743	1.291	1.6094	1.386	1.492	1.4917
3	priest	1.8971	1.8971	1.204	1.8971	0.9163	1.8971	1.743	1.6094	1.204	1.8971	1.8971	0.9808	1.4917	1.4917	1.609	1.6094	1.204	2.079	1.609	1.291
4	magistrate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	murder	0.8557	1.4917	0.8557	1.291	3.6889	0.8557	1.291	0.9808	1.743	0.8557	1.743	0.9808	1.204	1.3863	1.291	1.1239	1.204	2.303	1.386	1.291
6	day	1.8971	1.743	1.3863	1.743	1.0498	1.8971	2.303	1.6094	1.386	1.8971	1.743	1.1239	1.4917	1.743	1.609	1.8971	1.291	1.897	1.609	1.4917
7	embassy	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	police	1.204	1.3863	1.8971	1.3863	0.9163	1.204	1.386	1.3863	1.204	1.204	1.3863	0.9808	1.291	1.4917	1.386	1.8971	1.1239	1.492	1.897	2.0794
9	source	2.0794	1.4917	1.743	1.4917	0.9808	2.0794	2.079	1.4917	1.291	2.0794	1.4917	1.0498	2.0794	1.743	1.743	1.8971	1.743	1.743	1.897	1.4917
10	insult	1.291	1.6094	1.291	1.6094	1.204	1.291	2.079	1.4917	1.897	1.291	1.4917	1.204	1.4917	2.3026	1.897	1.743	1.743	1.609	1.492	1.6094
11	team	1.291	1.4917	1.8971	1.4917	0.9808	1.291	1.492	1.4917	1.291	1.291	1.4917	1.0498	1.3863	1.6094	1.492	2.0794	1.204	1.609	1.897	2.0794
12	claim	1.204	1.743	1.204	1.743	1.291	1.204	2.079	1.3863	2.079	1.204	1.3863	1.291	1.6094	2.5903	2.59	1.6094	1.8971	1.609	1.609	1.743
13	arrest	1.291	2.0794	1.291	3.6889	1.291	1.291	1.897	1.743	1.743	1.291	2.0794	1.291	1.6094	1.8971	1.743	1.743	1.6094	2.303	1.609	1.743
14	warrant	1.291	1.4917	1.291	1.4917	0.9808	1.291	2.079	1.4917	1.291	1.291	1.4917	1.0498	1.3863	2.3026	1.897	1.743	1.204	1.609	1.386	1.3863
15	attorney	1.8971	1.204	1.204	1.204	0.7985	1.8971	1.204	1.204	1.05	1.8971	1.204	0.8557	1.4917	1.291	1.609	1.3863	1.204	1.291	1.609	1.1239
16	envoy	1.743	1.204	1.1239	1.204	0.7985	1.743	1.386	1.204	1.05	1.8971	1.204	0.8557	1.3863	1.4917	1.492	1.3863	1.1239	1.291	1.492	1.1239
17	extradition	0.9163	1.291	0.9163	1.291	0.9808	0.9163	1.291	1.0498	1.291	0.9163	1.0498	3.6889	1.291	1.3863	1.291	1.204	1.204	1.291	1.204	1.4917
18	request	1.3863	1.8971	1.3863	1.8971	1.3863	1.3863	2.303	1.6094	1.897	1.3863	1.6094	1.3863	1.743	3.6889	2.996	1.8971	1.743	1.743	1.743	1.8971
19	opinion	1.3863	1.6094	1.3863	1.6094	1.204	1.3863	2.303	1.6094	1.609	1.3863	1.6094	1.3863	1.6094	2.5903	2.079	1.8971	1.4917	1.743	1.492	1.8971
20	term	1.3863	1.6094	1.3863	1.6094	1.0498	1.3863	2.303	1.6094	1.386	1.3863	1.6094	1.1239	1.4917	2.3026	2.303	1.8971	1.291	2.303	1.492	1.4917
21	rally	1.291	1.8971	1.743	1.8971	1.3863	1.291	1.897	1.4917	1.897	1.291	1.4917	1.6094	1.8971	2.0794	1.897	2.5903	1.743	1.897	2.303	2.3026
22	week	1.291	1.4917	1.291	1.4917	0.9808	1.291	2.303	1.4917	1.291	1.291	1.4917	1.0498	1.3863	1.6094	1.492	1.743	1.204	1.743	1.386	1.3863
23	boss	2.0794	1.291	1.8971	1.291	0.8557	2.0794	1.386	1.291	1.124	2.0794	1.291	0.9163	1.6094	1.3863	1.743	1.4917	1.8971	1.386	2.079	1.204
24	government	1.204	1.743	1.8971	1.743	1.291	1.204	1.743	1.3863	1.743	1.204	1.3863	1.4917	1.743	1.8971	1.743	1.8971	1.743	1.743	2.996	3.6889
25	official	2.0794	1.291	1.291	1.291	0.8557	2.0794	1.291	1.291	1.124	2.3026	1.291	0.9163	1.6094	1.3863	1.743	1.4917	1.291	1.386	1.743	1.204
26	post	1.8971	1.8971	1.6094	1.8971	1.3863	1.8971	2.079	1.6094	2.079	2.0794	1.6094	1.3863	1.743	2.0794	2.079	2.3026	1.8971	1.743	1.743	1.8971
27	crime	1.0498	1.6094	1.0498	1.6094	1.204	1.0498	1.897	1.204	1.897	1.0498	1.204	1.204	1.4917	1.743	1.897	1.3863	1.743	1.492	1.492	1.6094
28	prosecution	1.3863	1.743	1.6094	1.743	1.291	1.3863	2.079	1.6094	2.079	1.3863	1.6094	1.291	1.6094	1.8971	2.079	2.3026	1.8971	1.743	1.743	1.743
29	institution	1.3863	1.743	2.9957	1.4917	1.743	1.3863	2.303	1.4917	2.079	1.743	1.743	1.1239	1.4917	1.6094	1.492	2.0794	1.8971	1.897	2.079	2.3026
30	power	2.0794	2.3026	1.6094	2.3026	1.204	2.0794	2.303	2.3026	1.609	2.0794	2.3026	1.291	2.0794	1.743	2.303	1.8971	1.4917	2.59	1.743	2.0794
31	abuse	1.291	1.743	1.291	1.743	1.291	1.291	2.079	1.4917	2.079	1.291	1.4917	1.291	1.6094	2.3026	2.079	1.743	1.8971	1.609	1.609	1.743
32	constitution	1.3863	1.8971	1.3863	1.8971	1.743	1.3863	1.743	1.8971	2.079	1.3863	2.0794	1.1239	1.4917	1.743	1.609	1.8971	1.3863	2.079	1.609	1.4917
33	statement	1.3863	1.8971	1.3863	1.8971	1.3863	1.3863	2.303	1.6094	1.897	1.3863	1.6094	1.3863	1.743	2.5903	2.59	1.8971	1.743	1.743	1.743	1.8971
34	word	1.291	1.743	1.291	1.743	1.291	1.291	2.079	1.4917	1.743	1.291	1.4917	1.291	1.6094	2.3026	2.303	1.8971	1.6094	1.609	1.609	1.743
35	military	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
36	support	1.3863	1.8971	1.8971	1.8971	1.3863	1.3863	2.303	1.3863	2.303	1.743	1.3863	1.3863	1.743	2.0794	2.303	1.743	2.0794	1.743	2.079	1.8971
37	portrait	1.204	1.291	1.3863	1.291	0.8557	1.204	1.743	1.291	1.124	1.4917	1.291	0.9163	1.204	1.8971	1.897	1.4917	1.3863	1.386	1.492	1.204
38	extradition	0.9163	1.291	0.9163	1.291	0.9808	0.9163	1.291	1.0498	1.291	0.9163	1.0498	3.6889	1.291	1.3863	1.291	1.204	1.204	1.291	1.204	1.4917
39	law	1.3863	1.6094	1.8971	1.6094	1.1239	1.3863	2.996	1.6094	1.743	1.3863	1.6094	1.1239	1.4917	1.743	1.743	2.3026	1.6094	1.743	1.897	2.0794

40	protester	2.3026	1.3863	1.3863	1.3863	0.9163	2.3026	1.386	1.3863	1.204	2.3026	1.3863	0.9808	1.743	1.4917	1.897	1.6094	1.3863	1.492	1.897	1.291
41	arrest	1.291	2.0794	1.291	3.6889	1.291	1.291	1.897	1.743	1.743	1.291	2.0794	1.291	1.6094	1.8971	1.743	1.743	1.6094	2.303	1.609	1.743
42	opponent	2.3026	1.3863	1.3863	1.3863	0.9163	2.3026	1.386	1.3863	1.204	2.3026	1.3863	0.9808	1.743	1.4917	1.897	1.6094	1.3863	1.492	2.079	1.291
43	rightist	1.8971	1.204	1.204	1.204	0.7985	1.8971	1.204	1.204	1.05	1.8971	1.204	0.8557	1.4917	1.291	1.609	1.3863	1.204	1.291	1.609	1.1239
44	politician	2.0794	1.291	1.291	1.291	0.8557	2.0794	1.291	1.291	1.124	2.0794	1.291	0.9163	1.6094	1.3863	1.743	1.4917	1.291	1.386	1.743	1.204
45	time	1.4917	2.0794	1.4917	2.0794	1.204	1.4917	2.303	2.0794	1.609	1.4917	2.0794	1.291	2.0794	1.8971	1.743	2.0794	1.4917	2.303	1.609	1.743
46	amnesty	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
47	month	1.3863	1.6094	1.3863	1.6094	1.0498	1.3863	2.303	1.6094	1.386	1.3863	1.6094	1.1239	1.4917	1.743	1.609	1.8971	1.291	1.897	1.492	1.4917
48	father	2.0794	1.743	1.291	1.743	0.8557	2.0794	1.609	1.4917	1.124	2.0794	1.743	0.9163	1.6094	1.3863	1.743	1.4917	1.291	1.897	1.743	1.204
49	immunity	1.291	2.0794	1.291	2.0794	1.743	1.291	2.303	1.743	1.743	1.291	3.6889	1.0498	1.3863	1.6094	1.609	1.743	1.204	2.303	1.386	1.3863
50	caption	1.0498	1.6094	1.0498	1.6094	1.204	1.0498	1.609	1.204	1.609	1.0498	1.204	1.204	1.4917	2.0794	1.897	1.3863	1.4917	1.492	1.492	1.6094
51	capital	1.3863	1.4917	1.6094	1.4917	0.9808	1.3863	1.492	1.4917	1.291	1.3863	1.4917	1.0498	1.4917	1.6094	2.079	1.8971	1.291	1.609	2.303	2.5903
52	publicity	1.3863	1.8971	1.3863	1.8971	1.0498	1.3863	2.303	2.3026	1.386	1.3863	1.8971	1.1239	1.4917	2.5903	2.079	1.8971	1.291	2.079	1.492	1.4917
53	patient	2.0794	1.291	1.291	1.291	0.8557	2.0794	1.291	1.291	1.124	2.0794	1.291	0.9163	1.6094	1.3863	1.743	1.743	1.291	1.386	1.743	1.3863
54	arrest	1.291	2.0794	1.291	3.6889	1.291	1.291	1.897	1.743	1.743	1.291	2.0794	1.291	1.6094	1.8971	1.743	1.743	1.6094	2.303	1.609	1.743
55	order	1.3863	2.3026	1.8971	2.3026	1.291	1.3863	2.996	1.8971	2.079	1.3863	2.5903	1.291	1.6094	2.9957	2.59	2.3026	1.8971	2.59	1.897	2.0794
56	advice	1.291	1.4917	1.291	1.4917	0.9808	1.291	2.079	1.4917	1.291	1.291	1.4917	1.0498	1.3863	2.3026	1.897	1.743	1.204	1.609	1.386	1.3863
57	government	1.204	1.743	1.8971	1.743	1.291	1.204	1.743	1.3863	1.743	1.204	1.3863	1.4917	1.743	1.8971	1.743	1.8971	1.743	1.743	2.996	3.6889
58	lawmaker	2.0794	1.291	1.291	1.291	0.8557	2.0794	1.291	1.291	1.124	2.0794	1.291	0.9163	1.6094	1.3863	1.743	1.4917	1.291	1.386	1.743	1.204
59	bed	1.6094	1.291	1.8971	1.291	0.8557	1.6094	1.609	1.291	1.124	2.0794	1.291	0.9163	1.6094	1.3863	1.743	1.743	1.8971	1.386	2.079	1.204
60	result	1.6094	1.6094	1.291	1.6094	1.204	1.6094	2.079	1.4917	1.609	1.6094	1.4917	1.291	3.6889	2.3026	2.303	1.743	1.4917	2.079	1.492	1.743
61	water	1.743	1.4917	1.743	1.4917	0.9808	1.743	1.492	1.4917	1.291	1.8971	1.4917	1.0498	1.8971	1.6094	1.897	1.743	1.743	1.609	1.897	1.3863
62	cannon	1.4917	1.6094	1.3863	1.3863	1.3863	1.4917	1.743	1.291	2.303	1.4917	1.3863	1.0498	1.6094	1.4917	1.386	1.4917	1.3863	1.492	1.492	1.3863
63	decade	1.291	1.4917	1.291	1.4917	0.9808	1.291	2.303	1.4917	1.291	1.291	1.4917	1.0498	1.3863	1.6094	1.492	1.743	1.204	1.743	1.386	1.3863
64	application	1.4917	1.8971	1.291	1.8971	1.6094	1.4917	2.079	1.4917	2.303	1.4917	1.6094	1.3863	1.743	2.9957	2.079	1.743	1.8971	1.743	2.079	1.8971
65	regime	1.204	1.3863	1.8971	1.3863	0.9808	1.204	2.079	1.3863	1.291	1.204	1.3863	1.0498	1.3863	1.4917	1.386	1.8971	1.291	1.492	2.996	3.6889
66	delegation	1.291	1.4917	2.0794	1.4917	1.1239	1.291	1.492	1.4917	1.492	1.291	1.4917	1.291	1.4917	1.6094	1.492	2.0794	1.3863	1.609	2.079	2.3026
67	place	1.4917	2.0794	1.3863	2.0794	1.291	1.4917	2.303	1.743	2.079	1.4917	2.5903	1.291	1.743	1.8971	2.079	2.0794	1.8971	2.303	2.079	1.743
68	politician	2.0794	1.291	1.291	1.291	0.8557	2.0794	1.291	1.291	1.124	2.0794	1.291	0.9163	1.6094	1.3863	1.743	1.4917	1.291	1.386	1.743	1.204
69	scene	1.3863	2.0794	1.4917	2.0794	1.204	1.3863	1.897	1.743	1.609	1.6094	2.0794	1.291	2.0794	1.743	1.743	2.0794	1.4917	2.303	1.897	1.743
70	decision	1.291	1.743	1.291	1.743	1.4917	1.291	1.743	1.743	2.996	1.291	1.743	1.291	2.9957	1.8971	1.743	1.743	1.6094	1.897	1.897	1.743
71	judge	2.0794	1.291	1.291	1.291	0.8557	2.0794	1.291	1.291	1.124	2.0794	1.291	0.9163	1.6094	1.3863	1.743	1.4917	1.291	1.386	1.743	1.204
72	evening	1.291	1.4917	1.291	1.4917	0.9808	1.291	2.303	1.4917	1.291	1.291	1.4917	1.0498	1.3863	1.6094	1.492	1.743	1.204	1.743	1.386	1.3863
73	dictator	3.6889	1.291	1.291	1.291	0.8557	2.0794	1.291	1.291	1.124	2.0794	1.291	0.9163	1.6094	1.3863	1.743	1.4917	1.291	1.386	1.743	1.204
74	iron	1.3863	1.291	1.6094	1.291	0.8557	1.3863	1.609	1.291	1.124	1.743	1.291	0.9163	1.4917	1.3863	1.609	1.6094	1.6094	1.386	1.743	1.204
75	fist	1.204	1.0498	0.9808	1.0498	0.6931	1.204	1.05	1.0498	0.916	1.204	1.0498	0.7444	1.291	1.1239	1.05	1.204	0.9808	1.124	1.05	0.9808
76	visitor	2.0794	1.291	1.291	1.291	0.8557	2.0794	1.291	1.291	1.124	2.0794	1.291	0.9163	1.6094	1.3863	1.743	1.4917	1.291	1.386	1.743	1.204
77	diplomat	1.743	1.1239	1.1239	1.1239	0.7444	1.743	1.124	1.1239	0.981	2.0794	1.1239	0.7985	1.3863	1.204	1.492	1.291	1.1239	1.204	1.492	1.0498
78	policy	1.0498	1.204	1.0498	1.204	0.9163	1.0498	1.897	1.204	1.204	1.0498	1.204	0.9808	1.291	1.743	1.743	1.3863	1.204	1.291	1.124	1.3863
79	trial	1.204	1.743	1.204	1.743	1.291	1.204	2.079	1.3863	2.079	1.204	1.3863	1.3863	1.8971	1.8971	2.079	1.6094	1.8971	1.897	1.609	1.8971
80	black	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
81	police	1.204	1.3863	1.8971	1.3863	0.9163	1.204	1.386	1.3863	1.204	1.204	1.3863	0.9808	1.291	1.4917	1.386	1.8971	1.1239	1.492	1.897	2.0794
82	guard	1.8971	1.6094	1.743	1.6094	1.204	1.8971	1.897	1.6094	2.303	1.8971	1.743	1.204	1.4917	1.743	2.59	1.743	1.743	1.743	2.079	1.8971
83	stuff	1.743	1.8971	1.6094	1.8971	1.0498	1.743	2.079	2.3026	1.386	1.743	1.8971	1.1239	1.8971	2.3026	2.303	1.8971	1.6094	2.079	1.743	1.6094
84	idea	1.291	1.4917	1.291	1.4917	1.1239	1.291	2.59	1.4917	1.492	1.291	1.4917	1.204	1.6094	1.743	1.609	1.743	1.4917	1.609	1.386	1.743
85	business	1.3863	1.8971	1.8971	1.8971	1.3863	1.3863	2.303	1.6094	2.303	1.3863	1.6094	1.6094	1.743	2.0794	2.303	2.3026	2.0794	1.743	1.897	2.0794
86	army	1.204	1.3863	1.743	1.3863	0.9163	1.204	1.386	1.3863	1.204	1.204	1.3863	0.9808	1.291	1.4917	1.386	2.3026	1.1239	1.492	2.079	1.8971
87	people	1.4917	1.743	1.743	1.743	1.1239	1.4917	1.743	1.743	1.492	1.4917	1.743	1.204	1.6094	1.8971	1.743	2.5903	1.3863	1.897	1.897	1.8971
88	administration	1.291	1.743	1.743	1.743	1.291	1.291	1.897	1.4917	1.743	1.291	1.4917	1.4917	1.743	1.8971	1.743	2.0794	1.6094	1.743	1.897	3.6889
89	government	1.204	1.743	1.8971	1.743	1.291	1.204	1.743	1.3863	1.743	1.204	1.3863	1.4917	1.743	1.8971	1.743	1.8971	1.743	1.743	2.996	3.6889
90	report	1.291	1.743	1.291	1.743	1.291	1.291	2.079	1.4917	1.743	1.291	1.4917	1.291	1.8971	2.3026	2.079	1.743	1.6094	1.897	1.609	1.743
91	disappearance	1.291	1.8971	1.291	1.8971	2.0794	1.291	1.897	1.4917	2.079	1.291	2.0794	1.3863	2.3026	2.0794	1.897	1.743	1.743	2.303	1.743	1.8971
92	passport	1.291	1.743	1.291	1.743	0.9808	1.291	1.897	2.0794	1.291	1.291	1.743	1.0498	1.3863	2.0794	1.743	1.743	1.204	1.897	1.386	1.3863
93	family	2.0794	1.6094	1.8971	1.6094	1.0498	2.0794	1.609	1.6094	1.386	2.0794	1.6094	1.1239	1.6094	1.743	1.743	2.3026	1.291	1.743	1.897	2.0794
94	total	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
95	police	1.204	1.3863	1.8971	1.3863	0.9163	1.204	1.386	1.3863	1.204	1.204	1.3863	0.9808	1.291	1.4917	1.386	1.8971	1.1239	1.492	1.897	2.0794
96	gas	1.6094	1.8971	1.291	1.8971	0.9163	1.6094	2.079	1.6094	1.204	1.6094	2.3026	0.9808	1.8971	1.4917	1.743	1.6094	1.291	2.079	1.386	1.291
97	allegation	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
98	news	1.291	1.6094	1.291	1.6094	1.1239	1.291	2.079	1.8971	1.492	1.291	1.6094	1.204	1.6094	2.3026	1.897	1.743	1.3863	1.743	1.386	1.6094
99	court	1.743	1.6094	2.0794	1.6094	1.3863	1.743	1.609	1.3863	1.897	1.743	1.3863	1.204	1.4917	1.743	1.609	2.3026	2.5903	1.492	3.689	2.9957
100	strongman	2.9957	1.204	1.204	1.204	0.7985	1.8971	1.204	1.204	1.05	1.8971	1.204	0.8557	1.4917	1.291	1.609	1.3863	1.204	1.291	1.609	1.1239
101	rights	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

102	abuse	1.291	1.743	1.291	1.743	1.291	1.291	2.079	1.4917	2.079	1.291	1.4917	1.291	1.6094	2.3026	2.079	1.743	1.8971	1.609	1.609	1.743
103	country	1.4917	1.6094	1.743	1.6094	1.0498	1.4917	1.609	1.6094	1.386	1.4917	1.6094	1.1239	1.6094	1.743	1.609	2.3026	1.3863	1.743	2.079	1.8971
104	demonstrator	1.8971	1.204	1.204	1.204	0.7985	1.8971	1.204	1.204	1.05	2.0794	1.204	0.8557	1.4917	1.291	1.609	1.3863	1.204	1.291	1.609	1.1239
105	operation	1.743	2.0794	1.3863	2.0794	1.3863	1.743	2.303	1.743	2.303	1.743	2.0794	1.3863	2.3026	2.0794	2.59	1.8971	3.6889	2.303	1.743	1.8971
106	official	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
107	event	1.6094	2.0794	1.4917	2.0794	1.4917	1.6094	2.079	1.743	2.079	1.6094	2.3026	1.6094	3.6889	2.3026	2.303	2.0794	1.8971	2.303	1.897	2.3026
108	dissident	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
109	television	1.204	0.9808	1.3863	0.9808	0.6444	1.204	1.386	0.9808	0.856	1.4917	0.9808	0.6931	1.204	1.0498	1.291	1.204	1.3863	1.05	1.492	0.9163
110	interview	1.1239	1.4917	1.1239	1.4917	1.1239	1.1239	1.492	1.291	1.492	1.1239	1.291	1.1239	1.3863	2.3026	2.079	1.4917	1.3863	1.386	1.386	1.4917
111	division	1.3863	2.0794	1.743	1.6094	1.4917	1.3863	2.59	1.6094	2.079	1.3863	1.6094	1.204	1.4917	2.0794	1.897	2.3026	1.743	1.743	1.743	1.8971
112	riot	1.291	2.5903	1.291	2.0794	1.3863	1.291	1.897	1.743	1.897	1.291	2.0794	1.204	1.4917	2.0794	1.743	1.743	1.4917	2.303	1.743	1.6094
113	police	1.204	1.3863	1.8971	1.3863	0.9163	1.204	1.386	1.3863	1.204	1.204	1.3863	0.9808	1.291	1.4917	1.386	1.8971	1.1239	1.492	1.897	2.0794
114	police	1.204	1.3863	1.8971	1.3863	0.9163	1.204	1.386	1.3863	1.204	1.204	1.3863	0.9808	1.291	1.4917	1.386	1.8971	1.1239	1.492	1.897	2.0794
115	line	1.6094	1.8971	1.8971	1.8971	1.3863	1.6094	2.303	1.8971	2.303	2.0794	1.8971	1.3863	1.743	2.0794	2.303	2.0794	2.0794	2.079	2.079	1.8971
116	spokesman	1.8971	1.204	1.204	1.204	0.7985	1.8971	1.204	1.204	1.05	2.3026	1.204	0.8557	1.4917	1.291	1.609	1.3863	1.204	1.291	1.609	1.1239
117	response	1.4917	1.8971	1.291	1.8971	1.3863	1.4917	2.079	1.743	1.897	1.4917	1.8971	1.3863	2.9957	2.5903	2.303	1.743	1.743	1.897	1.743	1.8971
118	death	1.3863	2.3026	1.3863	2.3026	2.3026	1.3863	2.079	1.8971	1.897	1.3863	2.3026	1.291	2.0794	1.743	1.897	1.8971	1.4917	3.689	1.492	1.743
119	pressure	1.3863	1.743	1.1239	1.743	1.291	1.3863	1.743	1.743	1.743	1.3863	1.8971	1.291	2.0794	1.8971	2.303	1.4917	1.6094	2.079	1.609	1.743
120	release	1.6094	1.8971	1.6094	1.8971	2.0794	1.6094	2.303	1.3863	2.59	1.743	2.9957	1.3863	2.0794	2.0794	2.303	1.6094	2.0794	2.996	1.897	1.8971
121	injury	1.1239	1.743	1.1239	1.743	1.291	1.1239	2.079	1.3863	2.079	1.1239	1.8971	1.291	1.6094	1.8971	2.079	1.4917	1.8971	1.897	1.609	1.743
122	violation	1.1239	1.743	1.1239	1.743	1.291	1.1239	2.079	1.291	2.079	1.1239	1.291	1.291	1.6094	1.8971	2.079	1.4917	1.8971	1.609	1.609	1.743
123	attitude	1.3863	1.743	1.3863	1.743	1.204	1.3863	2.303	1.743	1.609	1.3863	1.8971	1.291	1.743	1.743	1.743	1.8971	1.4917	1.897	1.492	1.743
124	lobby	1.3863	1.6094	1.743	1.6094	1.0498	1.3863	1.609	1.6094	1.386	1.6094	1.6094	1.1239	1.4917	1.743	1.609	2.3026	2.5903	1.743	2.59	1.8971
125	group	1.6094	1.8971	1.8971	1.8971	1.204	1.6094	1.897	1.8971	1.609	1.6094	1.8971	1.291	1.743	2.0794	1.897	2.9957	1.4917	2.079	2.079	2.0794
126	placard	1.3863	1.6094	1.3863	1.6094	1.0498	1.3863	1.897	1.6094	1.386	1.3863	1.6094	1.1239	1.4917	2.0794	1.743	1.8971	1.291	1.743	1.492	1.4917
127	bearing	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
128	condition	1.3863	2.3026	1.3863	2.3026	1.291	1.3863	2.59	1.8971	2.079	1.3863	2.9957	1.291	1.6094	2.3026	2.303	1.8971	2.3026	2.59	1.609	1.743
129	son	1.8971	1.204	1.204	1.204	0.7985	1.8971	1.386	1.204	1.05	1.8971	1.204	0.8557	1.4917	1.291	1.609	1.3863	1.204	1.291	1.609	1.204
130	year	1.291	1.4917	1.743	1.4917	0.9808	1.291	2.303	1.4917	1.291	1.291	1.4917	1.0498	1.3863	1.6094	1.492	2.5903	1.204	1.743	2.303	1.8971
131	demonstration	1.3863	1.6094	1.3863	1.6094	1.204	1.3863	1.897	1.6094	1.743	1.3863	1.6094	1.3863	1.6094	2.0794	1.743	1.8971	1.6094	1.743	1.492	1.8971
132	anonymity	1.1239	1.743	1.1239	1.743	0.8557	1.1239	1.609	1.4917	1.124	1.1239	1.743	0.9163	1.204	1.3863	1.386	1.4917	1.0498	1.897	1.204	1.204
133	reign	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
134	hospital	1.291	1.291	3.6889	1.291	0.8557	1.291	1.291	1.291	1.124	1.6094	1.291	0.9163	1.291	1.3863	1.386	1.743	1.743	1.386	2.079	1.8971
135	protest	1.204	1.743	1.204	1.743	1.291	1.204	1.743	1.3863	1.743	1.204	1.3863	1.4917	1.743	2.3026	2.079	1.6094	1.6094	1.743	1.609	2.0794

C) Resnik measure

No	OC	CC1	CC2	CC3	CC4	CC5	CC6	CC7	CC8	CC9	CC10	CC11	CC12	CC13	CC14	CC15	CC16	CC17	CC18	CC19	CC20
		dictator	confusion	hospital	arrest	murder	citizen	rule	legality	move	supporter	immunity	extradition	outcome	request	charge	community	surgery	death	court	government
1	defense	1.3696	2.6044	3.9425	2.6044	2.6044	1.3696	4.038	0.7794	3.3826	2.4934	2.6044	4.5251	4.093	5.4967	7.8145	3.8937	3.9425	4.093	3.9425	4.5251
2	minister	1.9033	2.6044	1.3696	2.6044	2.6044	1.9033	3.383	0.7794	3.3826	4.4837	2.6044	2.6044	2.2541	2.6044	3.3826	1.1692	3.3826	2.6044	2.6044	2.6044
3	priest	1.9033	3.1688	1.3696	3.1688	0.7794	1.9033	3.169	2.3982	0.7794	1.9033	3.1688	0.7794	0.7794	0.7794	3.1688	3.1688	1.3696	3.1688	1.9033	0.7794
4	magistrate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	murder	0	4.5949	0.7794	2.6044	7.8339	0	2.604	0.7794	4.5949	0	6.883	2.6044	2.2541	2.6044	2.6044	0.7794	2.6044	7.4617	4.1005	2.6044
6	day	5.9193	3.1688	1.3696	3.1688	0.7794	1.9033	3.762	2.3982	0.7794	1.9033	3.1688	0.7794	0.7794	0.7794	3.1688	3.1688	1.3696	3.1688	1.9033	0.7794
7	embassy	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	police	0	0.7794	3.8937	0.7794	0.7794	0	0.779	0.7794	0.7794	0	0.7794	0.7794	0.7794	0.7794	0.7794	3.8937	0.7794	0.7794	3.8937	3.8937
9	source	5.9193	1.7798	2.4934	1.7798	1.7798	1.9033	4.058	0.7794	1.7798	2.4934	1.7798	1.7798	4.093	3.0718	4.093	3.0088	3.3927	4.093	3.0088	3.3927
10	insult	0	2.6044	0.7794	2.6044	2.6044	0	4.038	0.7794	3.3826	0	2.6044	2.6044	2.2541	4.0382	4.0382	0.7794	3.3826	2.6044	2.6044	2.6044
11	team	0	0.7794	3.8937	0.7794	0.7794	0	0.779	0.7794	0.7794	0	0.7794	0.7794	0.7794	0.7794	0.7794	3.8937	0.7794	0.7794	3.8937	3.8937
12	claim	0	2.6044	0.7794	2.6044	2.6044	0	4.555	0.7794	3.3826	0	2.6044	2.6044	2.2541	7.7584	7.5463	0.7794	3.3927	2.7753	2.6044	3.3927
13	arrest	0	3.1688	0.7794	10.667	2.6044	0	3.169	2.3982	2.6044	0	3.1688	2.6044	2.2541	2.6044	3.1688	3.1688	2.6044	3.1688	2.6044	2.6044
14	warrant	0	0.7794	0.7794	0.7794	0.7794	0	4.038	0.7794	0.7794	0	0.7794	0.7794	0.7794	4.0382	5.4027	3.1379	0.7794	0.7794	0.7794	0.7794
15	attorney	1.9033	0	1.3696	0	0	1.9033	1.37	0	0	1.9033	0	0	0.6144	0	1.9033	1.1692	1.3696	0.6144	1.9033	0
16	envoy	5.9193	0.7794	1.3696	0.7794	0.7794	1.9033	3.072	0.7794	0.7794	4.4837	0.7794	0.7794	0.7794	3.0718	3.0718	1.1692	1.3696	0.7794	1.9033	0.7794
17	extradition	0	2.6044	0.7794	2.6044	2.6044	0	2.604	0.7794	2.6044	0	2.6044	0	2.2541	2.6044	2.6044	0.7794	2.6044	2.6044	2.6044	4.5251
18	request	0	2.6044	0.7794	2.6044	2.6044	0	4.038	0.7794	2.6044	0	2.6044	2.6044	2.2541	7.7584	6.883	0.7794	2.6044	2.6044	2.6044	2.6044
19	opinion	0	2.6044	0.7794	2.6044	2.6044	0	4.058	0.7794	2.6044	0	2.6044	4.5251	2.2541	4.0382	4.0382	0.7794	3.3927	2.7753	2.6044	4.5251
20	term	1.3696	2.3982	2.8722	2.3982	1.7798	1.3696	4.555	2.3982	1.7798	2.4934	2.3982	1.7798	1.7798	4.0382	4.9623	3.1379	3.3927	6.131	2.8722	3.3927
21	rally	0.6144	4.1005	3.5267	2.6044	4.1005	0.6144	2.604	0.7794	4.1005	0.6144	4.1005	4.5251	4.093	2.6044	4.093	5.4468	2.6044	4.1005	5.4468	4.5251
22	week	0	0.7794	0.7794	0.7794	0.7794	0	3.762	0.7794	0.7794	0	0.7794	0.7794	0.7794	0.7794	0.7794	0.7794	0.7794	2.9496	0.7794	0.7794
23	boss	1.9033	0	3.9425	0	0	1.9033	2.493	0	0	2.4934	0	0	0.6144	0	2.4934	1.1692	3.9425	0.6144	3.9425	0
24	government	0	2.6044	3.8937	2.6044	2.6044	0	3.393	0.7794	2.6044	0	2.6044	4.5251	2.2541	2.6044	2.6044	3.8937	6.8241	5.1285	6.3813	11.7658
25	official	1.9033	0	1.3696	0	0	1.9033	1.37	0	0	4.4837	0	0	0.6144	0	1.9033	1.1692	1.3696	0.6144	1.9033	0
26	post	5.9193	4.5949	2.8722	2.6044	4.5949	1.9033	4.368	0.7794	4.5949	4.4837	4.5949	2.6044	2.2541	3.0718	3.3826	3.0088	3.3826	4.5949	4.1005	2.8722
27	crime	0	2.6044	0.7794	2.6044	2.6044	0	3.383	0.7794	3.3826	0	2.6044	2.6044	2.2541	2.6044	3.3826	0.7794	3.3826	2.6044	2.6044	2.6044
28	prosecution	0	2.6044	2.8722	2.6044	2.6044	0	3.383	0.7794	3.3826	0	2.6044	4.5251	2.2541	2.6044	3.3826	2.8722	3.3826	2.6044	2.8722	4.5251
29	institution	1.3696	4.5949	8.3646	2.6044	5.6088	1.3696	7.576	0.7794	4.5949	2.4934	5.6088	2.6044	2.2541	2.6044	2.6044	3.8937	3.9425	5.6088	5.3823	3.8937
30	power	1.9033	3.1688	3.8937	3.1688	1.7798	1.9033	3.169	4.1033	1.7798	1.9033	7.6882	1.7798	4.5024	3.0718	4.8762	3.8937	2.7753	5.1285	3.8937	5.1285
31	abuse	0	2.6044	0.7794	2.6044	2.6044	0	4.038	0.7794	3.3826	0	2.6044	2.6044	2.2541	4.0382	4.0382	0.7794	3.3826	2.6044	2.6044	2.6044
32	constitution	1.3696	4.5949	2.4934	2.6044	5.6088	1.3696	3.445	2.3982	4.5949	2.4934	5.6088	2.6044	2.2541	3.0718	3.0718	2.3982	2.6044	5.6088	4.1005	2.6044
33	statement	0	2.6044	0.7794	2.6044	2.6044	0	4.038	0.7794	2.6044	0	2.6044	2.6044	2.2541	5.4967	5.4967	0.7794	2.6044	2.6044	2.6044	2.6044
34	word	0	2.6044	0.7794	2.6044	2.6044	0	4.038	0.7794	2.6044	0	2.6044	2.6044	2.2541	5.4967	7.7404	3.1379	3.3927	2.9496	2.6044	3.3927
35	military	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

36	support	1.3696	2.6044	3.9425	2.6044	2.6044	1.3696	4.368	0.7794	3.3826	2.4934	2.6044	2.6044	2.2541	3.0718	6.7685	3.1379	3.9425	2.7753	3.9425	2.7753
37	portrait	1.3696	0.7794	2.4934	0.7794	0.7794	1.3696	4.038	0.7794	0.7794	2.4934	0.7794	0.7794	0.7794	4.0382	4.9623	1.1692	2.4934	0.7794	2.4934	0.7794
38	extradition	0	2.6044	0.7794	2.6044	2.6044	0	2.604	0.7794	2.6044	0	2.6044	0	2.2541	2.6044	2.6044	0.7794	2.6044	2.6044	2.6044	4.5251
39	law	0	2.6044	3.8937	2.6044	2.6044	0	9.368	0.7794	3.3826	0	2.6044	2.6044	2.2541	3.0718	3.3826	3.8937	5.823	2.7753	3.8937	5.823
40	protester	1.9033	0	1.3696	0	0	1.9033	1.37	0	0	1.9033	0	0	0.6144	0	1.9033	1.1692	1.3696	0.6144	1.9033	0
41	arrest	0	3.1688	0.7794	10.667	2.6044	0	3.169	2.3982	2.6044	0	3.1688	2.6044	2.2541	2.6044	3.1688	3.1688	2.6044	3.1688	2.6044	2.6044
42	opponent	1.9033	0	1.3696	0	0	1.9033	1.37	0	0	1.9033	0	0	0.6144	0	1.9033	1.1692	1.3696	0.6144	5.9912	0
43	rightist	1.9033	0	1.3696	0	0	1.9033	1.37	0	0	1.9033	0	0	0.6144	0	1.9033	1.1692	1.3696	0.6144	1.9033	0
44	politician	1.9033	0	1.3696	0	0	1.9033	1.37	0	0	1.9033	0	0	0.6144	0	1.9033	1.1692	1.3696	0.6144	1.9033	0
45	time	0	2.3982	0.7794	2.3982	2.2541	0	3.762	2.3982	2.2541	0	3.8895	2.2541	4.0796	2.2541	2.3982	2.3982	2.7753	6.131	2.2541	2.7753
46	amnesty	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
47	month	0	0.7794	0.7794	0.7794	0.7794	0	3.762	0.7794	0.7794	0	0.7794	0.7794	0.7794	0.7794	0.7794	0.7794	0.7794	2.9496	0.7794	0.7794
48	father	1.9033	3.1688	1.3696	3.1688	1.7798	1.9033	3.393	2.3982	1.7798	1.9033	3.1688	1.7798	1.7798	1.7798	3.6967	3.1688	3.3927	3.1688	1.9033	3.3927
49	immunity	0	4.5949	0.7794	3.1688	6.883	0	4.143	2.3982	4.5949	0	9.2008	2.6044	2.2541	2.6044	3.1688	3.1688	2.6044	6.883	4.1005	2.6044
50	caption	0	2.6044	0.7794	2.6044	2.6044	0	3.072	0.7794	2.6044	0	2.6044	2.6044	2.2541	5.4967	5.4967	0.7794	2.6044	2.6044	2.6044	2.6044
51	capital	1.3696	0.7794	3.8937	0.7794	0.7794	1.3696	3.072	0.7794	0.7794	2.4934	0.7794	0.7794	0.7794	3.0718	4.6453	3.8937	2.4934	0.7794	6.3813	6.3813
52	publicity	0	2.3982	0.7794	2.3982	0.7794	0	4.038	4.1033	0.7794	0	2.3982	0.7794	0.7794	4.0382	4.0382	2.3982	0.7794	2.3982	0.7794	0.7794
53	patient	1.9033	0.7794	2.8722	0.7794	0.7794	1.9033	1.37	0.7794	0.7794	1.9033	0.7794	0.7794	0.7794	0.7794	1.9033	2.8722	1.3696	0.7794	2.8722	2.8722
54	arrest	0	3.1688	0.7794	10.667	2.6044	0	3.169	2.3982	2.6044	0	3.1688	2.6044	2.2541	2.6044	3.1688	3.1688	2.6044	3.1688	2.6044	2.6044
55	order	0	4.1426	5.6152	3.1688	2.6044	0	8.269	2.3982	3.3826	0	4.1426	2.6044	2.2541	6.883	7.7404	8.0769	3.3826	3.1688	3.8937	3.8937
56	advice	0	0.7794	0.7794	0.7794	0.7794	0	4.038	0.7794	0.7794	0	0.7794	0.7794	0.7794	4.0382	4.0382	0.7794	0.7794	0.7794	0.7794	0.7794
57	government	0	2.6044	3.8937	2.6044	2.6044	0	3.393	0.7794	2.6044	0	2.6044	4.5251	2.2541	2.6044	2.6044	3.8937	6.8241	5.1285	6.3813	11.7658
58	lawmaker	1.9033	0	1.3696	0	0	1.9033	1.37	0	0	1.9033	0	0	0.6144	0	1.9033	1.1692	1.3696	0.6144	1.9033	0
59	bed	1.3696	0	3.9425	0	0	1.3696	3.445	0	0	2.4934	0	0	0.6144	0	2.4934	3.6252	3.9425	0.6144	5.8659	0
60	result	0.6144	2.2541	2.8722	2.2541	2.2541	0.6144	4.038	0.7794	2.2541	0.6144	2.2541	2.2541	7.7053	4.0382	4.9623	2.8722	2.2541	4.5024	2.8722	2.8722
61	water	1.3696	0.7794	2.4934	0.7794	0.7794	1.3696	2.493	0.7794	0.7794	2.4934	0.7794	0.7794	0.7794	0.7794	5.3672	3.1379	2.4934	0.7794	2.4934	0.7794
62	cannon	1.3696	4.5949	2.4934	2.6044	4.5949	1.3696	5.436	0.7794	5.9546	5.2675	4.5949	2.6044	2.2541	2.6044	2.6044	1.1692	2.6044	4.5949	4.1005	2.6044

63	decade	0	0.7794	0.7794	0.7794	0.7794	0	3.762	0.7794	0.7794	0	0.7794	0.7794	0.7794	0.7794	0.7794	0.7794	0.7794	2.9496	0.7794	0.7794
64	application	1.8747	4.1005	0.7794	2.6044	4.1005	1.8747	4.038	0.7794	4.1005	1.8747	4.1005	2.6044	2.2541	7.7584	4.7636	0.7794	3.3826	4.1005	4.1005	2.6044
65	regime	0	1.7798	3.8937	1.7798	1.7798	0	4.058	0.7794	1.7798	0	1.7798	1.7798	1.7798	1.7798	1.7798	3.8937	3.3927	2.7753	6.3813	6.3813
66	delegation	0	2.6044	3.8937	2.6044	2.6044	0	2.604	0.7794	2.6044	0	2.6044	4.5251	2.2541	2.6044	2.6044	3.8937	2.6044	2.6044	3.8937	6.1239
67	place	1.1692	4.1426	1.1692	3.1688	2.6044	1.1692	4.143	2.3982	3.3826	1.1692	4.1426	2.6044	2.2541	3.0718	6.6659	3.6252	3.3826	3.1688	5.8659	2.7753
68	politician	1.9033	0	1.3696	0	0	1.9033	1.37	0	0	1.9033	0	0	0.6144	0	1.9033	1.1692	1.3696	0.6144	1.9033	0
69	scene	1.3696	4.6277	2.4934	3.1688	2.2541	1.3696	3.393	2.3982	2.2541	2.4934	3.1688	2.2541	4.0796	3.0718	5.5592	3.6252	3.3927	4.0796	4.679	3.3927
70	decision	0	4.1005	0.7794	2.6044	4.1005	0	3.393	2.3982	7.1706	0	4.1005	2.6044	7.7053	2.6044	2.6044	2.3982	3.3927	4.1005	4.1005	3.3927
71	judge	1.9033	0	1.3696	0	0	1.9033	1.37	0	0	4.4837	0	0	0.6144	0	1.9033	1.1692	1.3696	0.6144	1.9033	0
72	evening	0	0.7794	0.7794	0.7794	0.7794	0	3.762	0.7794	0.7794	0	0.7794	0.7794	0.7794	0.7794	0.7794	0.7794	0.7794	2.9496	0.7794	0.7794
73	dictator	10.38	0	1.3696	0	0	1.9033	1.37	0	0	1.9033	0	0	0.6144	0	1.9033	1.1692	1.3696	0.6144	1.9033	0
74	iron	1.3696	0.7794	2.4934	0.7794	0.7794	1.3696	3.445	0.7794	0.7794	5.8125	0.7794	0.7794	0.7794	0.7794	3.9759	3.1379	2.4934	0.7794	2.4934	0.7794
75	fist	0.6144	0	0.6144	0	0	0.6144	0.614	0	0	0.6144	0	0	0.6144	0	0.6144	0.6144	0.6144	0.6144	0.6144	0
76	visitor	1.9033	0	1.3696	0	0	1.9033	1.37	0	0	1.9033	0	0	0.6144	0	1.9033	1.1692	1.3696	0.6144	1.9033	0
77	diplomat	5.9193	0	1.3696	0	0	1.9033	1.37	0	0	4.4837	0	0	0.6144	0	1.9033	1.1692	1.3696	0.6144	1.9033	0
78	policy	0	1.7798	0.7794	1.7798	1.7798	0	4.058	0.7794	1.7798	0	1.7798	1.7798	1.7798	4.0382	4.9623	0.7794	3.3927	2.7753	1.7798	3.3927
79	trial	0	2.6044	0.7794	2.6044	2.6044	0	3.383	0.7794	3.3826	0	2.6044	4.5251	4.0796	2.6044	3.3826	0.7794	3.3826	4.0796	2.6044	4.5251
80	black	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
81	police	0	0.7794	3.8937	0.7794	0.7794	0	0.779	0.7794	0.7794	0	0.7794	0.7794	0.7794	0.7794	0.7794	3.8937	0.7794	0.7794	3.8937	3.8937
82	guard	1.9033	4.1005	3.8937	2.6044	4.1005	1.9033	4.368	2.3982	7.5914	2.4934	4.1005	2.6044	2.2541	2.6044	8.8754	3.8937	3.3826	4.1005	6.8531	3.8937
83	stuff	1.1692	2.3982	1.1692	2.3982	1.7798	1.1692	4.038	4.1033	1.7798	1.1692	2.3982	1.7798	1.7798	4.0382	4.6453	3.1379	3.3927	2.7753	1.7798	3.3927
84	idea	0	1.7798	0.7794	1.7798	1.7798	0	4.058	0.7794	1.7798	0	1.7798	1.7798	1.7798	3.0718	3.0718	0.7794	3.3927	2.7753	1.7798	3.3927
85	business	0	2.6044	3.8937	2.6044	2.6044	0	3.393	0.7794	3.3826	0	2.6044	5.985	2.2541	2.6044	4.7636	3.8937	5.823	2.7753	3.8937	5.823
86	army	0	0.7794	3.8937	0.7794	0.7794	0	0.779	0.7794	0.7794	0	0.7794	0.7794	0.7794	0.7794	0.7794	5.4468	0.7794	0.7794	5.4468	3.8937
87	people	0	0.7794	3.5267	0.7794	0.7794	0	0.779	0.7794	0.7794	0	0.7794	0.7794	0.7794	0.7794	0.7794	3.5267	0.7794	0.7794	3.5267	3.5267
88	administration	0	2.6044	3.5267	2.6044	2.6044	0	3.762	0.7794	3.3826	0	2.6044	4.5251	2.2541	2.6044	4.7636	3.5267	3.3826	2.9496	3.5267	9.1267
89	government	0	2.6044	3.8937	2.6044	2.6044	0	3.393	0.7794	2.6044	0	2.6044	4.5251	2.2541	2.6044	2.6044	3.8937	6.8241	5.1285	6.3813	11.7658
90	report	0	2.6044	0.7794	2.6044	2.6044	0	4.038	0.7794	2.6044	0	2.6044	2.6044	4.0796	5.4967	5.4967	0.7794	2.7753	4.0796	2.6044	2.7753
91	disappearance	0	4.5949	0.7794	2.6044	6.883	0	2.604	0.7794	4.5949	0	6.883	2.6044	4.0796	2.6044	2.6044	0.7794	2.6044	6.883	4.1005	2.6044
92	passport	0	2.3982	0.7794	2.3982	0.7794	0	4.038	4.1033	0.7794	0	2.3982	0.7794	0.7794	4.0382	4.0382	2.3982	0.7794	2.3982	0.7794	0.7794
93	family	1.9033	0.7794	3.8937	0.7794	0.7794	1.9033	1.37	0.7794	0.7794	1.9033	0.7794	0.7794	0.7794	0.7794	1.9033	3.8937	1.3696	0.7794	3.8937	3.8937
94	total	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
95	police	0	0.7794	3.8937	0.7794	0.7794	0	0.779	0.7794	0.7794	0	0.7794	0.7794	0.7794	0.7794	0.7794	3.8937	0.7794	0.7794	3.8937	3.8937

96	gas	1.3696	4.1426	2.4934	3.1688	0.7794	1.3696	4.143	2.3982	0.7794	2.4934	4.1426	0.7794	4.5024	0.7794	5.3672	3.1688	2.4934	4.7384	2.4934	0.7794
97	allegation	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
98	news	0	2.3982	0.7794	2.3982	2.2541	0	4.038	4.1033	2.2541	0	2.3982	2.2541	2.2541	4.0382	4.0382	2.3982	2.2541	2.3982	2.2541	2.2541
99	court	1.9033	4.1005	5.3823	2.6044	4.1005	1.9033	2.604	0.7794	4.1005	2.4934	4.1005	2.6044	2.2541	2.6044	2.6044	5.4468	5.8073	4.1005	11.073	6.3813
100	strongman	10.38	0	1.3696	0	0	1.9033	1.37	0	0	1.9033	0	0	0.6144	0	1.9033	1.1692	1.3696	0.6144	1.9033	0
101	rights	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
102	abuse	0	2.6044	0.7794	2.6044	2.6044	0	4.038	0.7794	3.3826	0	2.6044	2.6044	2.2541	4.0382	4.0382	0.7794	3.3826	2.6044	2.6044	2.6044
103	country	1.1692	0.7794	3.8937	0.7794	0.7794	1.1692	1.169	0.7794	0.7794	1.1692	0.7794	0.7794	0.7794	0.7794	1.1692	4.4243	1.1692	0.7794	4.679	3.8937
104	demonstrator	1.9033	0	1.3696	0	0	1.9033	1.37	0	0	4.4837	0	0	0.6144	0	1.9033	1.1692	1.3696	0.6144	1.9033	0
105	operation	0.6144	3.1688	0.7794	3.1688	2.6044	0.6144	6.824	2.3982	3.3826	0.6144	3.1688	2.6044	4.093	2.6044	6.7685	3.1688	10.38	4.093	2.6044	2.7753
106	official	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
107	event	0.6144	4.1426	0.7794	3.1688	2.2541	0.6144	4.143	2.3982	2.2541	0.6144	4.1426	2.2541	6.2403	2.2541	4.8762	3.1688	2.2541	4.7384	2.2541	2.2541
108	dissident	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
109	television	1.3696	0	2.4934	0	0	1.3696	3.445	0	0	2.4934	0	0	0.6144	0	2.4934	1.1692	2.4934	0.6144	2.4934	0
110	interview	0	2.6044	0.7794	2.6044	2.6044	0	3.072	0.7794	2.6044	0	2.6044	2.6044	2.2541	6.883	6.883	0.7794	2.6044	2.6044	2.6044	2.6044
111	division	0	7.131	3.8937	2.6044	4.5949	0	6.824	0.7794	4.5949	0	4.5949	2.6044	2.2541	5.4967	5.4967	3.8937	6.8241	4.5949	4.1005	3.8937
112	riot	0	7.6226	0.7794	3.1688	4.1005	0	4.038	2.3982	4.1005	0	4.1005	2.6044	2.2541	4.0382	4.0382	3.1688	3.3826	4.1005	4.1005	2.6044
113	police	0	0.7794	3.8937	0.7794	0.7794	0	0.779	0.7794	0.7794	0	0.7794	0.7794	0.7794	0.7794	0.7794	3.8937	0.7794	0.7794	3.8937	3.8937
114	police	0	0.7794	3.8937	0.7794	0.7794	0	0.779	0.7794	0.7794	0	0.7794	0.7794	0.7794	0.7794	0.7794	3.8937	0.7794	0.7794	3.8937	3.8937
115	line	1.3696	2.6044	3.8937	2.6044	2.6044	1.3696	4.368	2.3982	3.3826	5.5837	2.6044	4.5251	4.5024	3.0718	4.8762	3.8937	3.3826	4.7384	3.8937	4.5251
116	spokesman	1.9033	0	1.3696	0	0	1.9033	1.37	0	0	7.6549	0	0	0.6144	0	1.9033	1.1692	1.3696	0.6144	1.9033	0
117	response	0.6144	2.6044	2.8722	2.6044	2.6044	0.6144	4.038	2.3982	2.6044	0.6144	3.8895	2.6044	6.2403	5.4967	5.4967	2.8722	2.6044	4.5024	2.8722	2.8722
118	death	0.6144	4.5949	0.7794	3.1688	7.4617	0.6144	3.169	2.3982	4.5949	0.6144	6.883	2.6044	4.5024	2.6044	4.7384	3.1688	2.7753	10.6671	4.1005	5.1285
119	pressure	0.6144	4.1426	0.7794	3.1688	2.6044	0.6144	4.143	4.1033	2.6044	0.6144	4.1426	2.6044	4.5024	2.6044	4.8762	3.1688	2.7753	4.7384	2.6044	2.7753
120	release	1.3696	4.5949	2.4934	2.6044	6.883	1.3696	4.368	0.7794	5.9546	5.5837	6.883	2.6044	4.093	4.0382	4.9623	1.1692	3.3826	7.7955	4.1005	2.6044
121	injury	0	4.1426	0.7794	3.1688	2.6044	0	4.143	2.3982	3.3826	0	4.1426	2.6044	4.0796	2.6044	3.3826	3.1688	3.3826	5.8908	2.6044	2.6044
122	violation	0	2.6044	0.7794	2.6044	2.6044	0	3.383	0.7794	3.3826	0	2.6044	2.6044	2.2541	2.6044	3.3826	0.7794	3.3826	2.6044	2.6044	2.6044
123	attitude	0	2.6044	0.7794	2.6044	2.6044	0	3.383	2.3982	3.3826	0	3.8895	2.6044	2.2541	2.6044	3.3826	3.1379	3.3826	2.7753	2.6044	2.7753
124	lobby	1.3696	0.7794	3.9425	0.7794	0.7794	1.3696	2.493	0.7794	0.7794	2.4934	0.7794	0.7794	0.7794	0.7794	2.4934	3.8937	5.8073	0.7794	5.8073	3.8937
125	group	0.6144	0.7794	2.8722	0.7794	0.7794	0.6144	0.779	0.7794	0.7794	0.6144	0.7794	0.7794	0.7794	0.7794	0.7794	2.8722	0.7794	0.7794	2.8722	2.8722

126	placard	0	0.7794	0.7794	0.7794	0.7794	0	3.072	0.7794	0.7794	0	0.7794	0.7794	0.7794	3.0718	3.0718	0.7794	0.7794	0.7794	0.7794	0.7794
127	bearing	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
128	condition	0	4.1426	0.7794	3.1688	2.6044	0	6.824	2.3982	3.3826	0	4.1426	2.6044	2.2541	4.0382	4.9623	3.1688	6.8241	3.1688	2.6044	2.7753
129	son	1.9033	1.7798	1.3696	1.7798	1.7798	1.9033	3.393	0.7794	1.7798	1.9033	1.7798	1.7798	1.7798	1.7798	1.9033	1.1692	3.3927	2.7753	1.9033	3.3927
130	year	0	0.7794	3.5267	0.7794	0.7794	0	3.762	0.7794	0.7794	0	0.7794	0.7794	0.7794	0.7794	0.7794	5.4468	0.7794	2.9496	5.4468	3.5267
131	demonstration	0	2.6044	0.7794	2.6044	2.6044	0	3.383	0.7794	3.3826	0	2.6044	4.5251	2.2541	3.0718	3.3826	0.7794	3.3826	2.6044	2.6044	4.5251
132	anonymity	0	3.1688	0.7794	3.1688	0.7794	0	3.169	2.3982	0.7794	0	3.1688	0.7794	0.7794	0.7794	3.1688	3.1688	0.7794	3.1688	0.7794	0.7794
133	reign	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
134	hospital	1.3696	0.7794	8.3646	0.7794	0.7794	1.3696	2.493	0.7794	0.7794	2.4934	0.7794	0.7794	0.7794	0.7794	2.4934	3.8937	3.9425	0.7794	5.3823	3.8937
135	protest	0	2.6044	0.7794	2.6044	2.6044	0	2.604	0.7794	2.6044	0	2.6044	4.5251	2.2541	5.4967	5.4967	0.7794	2.6044	2.6044	2.6044	4.5251

D) LIN measure

No	OC	CC1	CC2	CC3	CC4	CC5	CC6	CC7	CC8	CC9	CC10	CC11	CC12	CC13	CC14	CC15	CC16	CC17	CC18	CC19	CC20
		dictator	confusion	hospital	arrest	murder	citizen	rule	legality	move	supporter	immunity	extradition	outcome	request	charge	community	surgery	death	court	government
1	defense	0.1404	0.0923	0.4508	0.2737	0.3117	0.1599	0.4661	0.0781	0.3512	0.1638	0.0862	0	0.4728	0.5895	0.7924	0.4209	0.3391	0.4286	0.4089	0.5027
2	minister	0.1993	0	0.1603	0	0	0.2276	0.1337	0	0	0.2334	0	0	0.0821	0	0.1858	0.1341	0	0.0734	0.1451	0
3	priest	0.2023	0.3855	0.1631	0.3318	0.0958	0.2316	0.3369	0.2459	0.1083	0.2375	0.3594	0	0.0966	0.1018	0.3443	0.3138	0.0829	0.3732	0.1473	0.1052
4	magistrate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5	murder	0	0.0984	0	0.2895	1	0	0.2211	0.0824	0.6665	0	0.0915	0	0.2901	0.3539	0.3095	0.1055	0.286	0.2884	0.0991	0.3071
6	day	0	0.1095	0	0.0951	0.1108	0	0.1075	0.0901	0.1279	0	0.101	0	0.1118	0.1188	0.1141	0.1183	0.0938	0.3931	0.1103	0.1235
7	embassy	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8	police	0	0.0982	0	0.0865	0.0992	0	0.0966	0.0823	0.1127	0	0.0913	0	0.1001	0.1056	0.1019	0.4762	0.0854	0.0995	0.4473	0.5463
9	source	0.1885	0.0906	0.2763	0.1839	0.209	0.2136	0.4583	0.0769	0.2349	0.2186	0.0847	0	0.2106	0.3623	0.3498	0.3655	0.1818	0.2794	0.3269	0.3505
10	insult	0	0.0817	0	0.2454	0.2755	0	0.4176	0.0704	0.3059	0	0.0769	0	0.2401	0.4289	0.4155	0.0866	0.3154	0.2389	0.0822	0.2579
11	team	0	0.1009	0	0.0885	0.102	0	0.0992	0.0842	0.1163	0	0.0936	0	0.1029	0.1088	0.1048	0.4903	0.0874	0.1023	0.4597	0.5631
12	claim	0	0.0944	0	0.2608	0.2951	0	0.4814	0.0796	0.3302	0	0.088	0	0.2572	0.824	0.5882	0.1009	0.3349	0.2602	0.095	0.3259
13	arrest	0	0.3394	0	1	0.2895	0	0.3011	0.2206	0.3233	0	0.319	0	0.2524	0.3057	0.307	0.2825	0.2536	0.3298	0.0864	0.2701
14	warrant	0	0.0848	0	0.0759	0.0856	0	0.4331	0.0727	0.0954	0	0.0796	0	0.0862	0.4453	0.4309	0.09	0.0751	0.0858	0.0853	0.093
15	attorney	0.206	0	0.1663	0	0	0.2363	0.1379	0	0	0.2426	0	0	0.0857	0	0.1916	0.139	0	0.0762	0.15	0
16	envoy	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	extradition	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	request	0	0.1047	0	0.3057	0.3539	0	0.5039	0.0868	0.4057	0	0.0969	0	0.309	1	0.5009	0.1128	0.3017	0.3071	0.1055	0.3254
19	opinion	0	0.1048	0	0.2588	0.2925	0	0.5279	0.0869	0.327	0	0.097	0	0.255	0.4783	0.4617	0.1129	0.2559	0.3165	0.1056	0.4738
20	term	0	0.0965	0	0.1709	0.1924	0	0.4751	0.0811	0.2142	0	0.0898	0	0.1937	0.4555	0.5412	0.3386	0.1691	0.339	0.2922	0.3259
21	rally	0	0.0817	0	0.2376	0.4184	0	0.1777	0.0704	0.4628	0	0.0769	0	0.2315	0.2793	0.2509	0.6049	0.2352	0.2305	0.5743	0.4041
22	week	0	0.1031	0	0.0902	0.1042	0	0.1013	0.0857	0.1192	0	0.0955	0	0.1051	0.1113	0.1072	0.1109	0.0891	0.3712	0.1038	0.1154
23	boss	0.1908	0	0.4057	0	0	0.2166	0.2184	0	0	0.2218	0	0	0.0777	0	0.1784	0.1279	0	0.0698	0.3714	0
24	government	0	0.1084	0	0.2701	0.3071	0	0.3636	0.0893	0.3454	0	0.1	0	0.2678	0.3254	0.2875	0.5296	0.267	0.4572	0.7486	1
25	official	0.2191	0	0.1783	0	0	0.2538	0.146	0	0	0.6148	0	0	0.0928	0	0.2029	0.1488	0	0.0818	0.1597	0
26	post	0.1373	0.0983	0.2781	0.2892	0.332	0.1559	0.4094	0.0824	0.3772	0.1596	0.0914	0	0.2898	0.3535	0.4016	0.315	0.371	0.2881	0.2859	0.3068
27	crime	0	0.0993	0	0.2919	0.3356	0	0.2231	0.0831	0.3818	0	0.0923	0	0.2929	0.3575	0.4055	0.1066	0.3744	0.2912	0.1	0.3098
28	prosecution	0	0.0835	0	0.2501	0.2815	0	0.188	0.0717	0.3134	0	0.0785	0	0.2454	0.2968	0.2649	0.3071	0.2475	0.2442	0.2922	0.4572
29	institution	0.1441	0.1145	0.879	0.2588	0.6299	0.1647	0.7563	0.0934	0.5769	0.1689	0.1052	0	0.255	0.309	0.2746	0.5619	0.2559	0.2537	0.5221	0.6491
30	power	0.1869	0.3606	0.1439	0.3132	0.2746	0.2119	0.4143	0.4744	0.3212	0.217	0.3377	0	0.5886	0.3638	0.5076	0.4007	0.2295	0.6494	0.38	0.6071
31	abuse	0	0.0906	0	0.2691	0.3058	0	0.4176	0.0769	0.3437	0	0.0847	0	0.2667	0.4289	0.4155	0.0966	0.3455	0.2652	0.0912	0.2842
32	constitution	0	0.2851	0	0.2729	0.669	0	0.3334	0.2411	0.6173	0	0.2661	0	0.271	0.3429	0.3317	0.233	0.2697	0.2762	0.0933	0.2884
33	statement	0	0.1202	0	0.2376	0.2658	0	0.6104	0.0972	0.2939	0	0.1101	0	0.2315	0.6349	0.7447	0.131	0.2352	0.2305	0.1212	0.2493
34	word	0	0.1112	0	0.2501	0.2815	0	0.4846	0.0912	0.3134	0	0.1024	0	0.2454	0.6264	0.5757	0.3827	0.2475	0.2949	0.112	0.3706
35	military	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
36	support	0.1565	0.0983	0.4397	0.2892	0.332	0.1811	0.4625	0.0824	0.3772	0.1862	0.0914	0	0.2898	0.3535	0.7247	0.3205	0.371	0.2881	0.3998	0.3068
37	portrait	0.1277	0.0835	0.2566	0.0749	0.0843	0.1436	0.4265	0.0717	0.0938	0.1468	0.0785	0	0.0848	0.4383	0.5215	0.1181	0.0741	0.0844	0.2349	0.0914
38	extradition	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	law	0	0.102	0	0.2454	0.2755	0	0.9351	0.085	0.3059	0	0.0946	0	0.2401	0.3873	0.3731	0.4762	0.3154	0.29	0.4473	0.5733
40	protester	0.1885	0	0.1506	0	0	0.2136	0.1269	0	0	0.2186	0	0	0.0765	0	0.1763	0.1261	0	0.0688	0.1371	0

41	arrest	0	0.3394	0	1	0.2895	0	0.3011	0.2206	0.3233	0	0.319	0	0.2524	0.3057	0.307	0.2825	0.2536	0.3298	0.0864	0.2701
42	opponent	0.206	0	0.1663	0	0	0.2363	0.1379	0	0	0.2426	0	0	0.0857	0	0.1916	0.139	0	0.0762	0.15	0
43	rightist	0.1908	0	0.1527	0	0	0.2166	0.1284	0	0	0.2218	0	0	0.0777	0	0.1784	0.1279	0	0.0698	0.1389	0
44	politician	0.2111	0	0.171	0	0	0.2431	0.141	0	0	0.2497	0	0	0.0884	0	0.196	0.1428	0	0.0783	0.1538	0
45	time	0	0.343	0	0.2881	0.3173	0	0.3546	0.2813	0.3657	0	0.316	0	0.5796	0.3401	0.3007	0.2703	0.2691	0.732	0.1124	0.3165
46	amnesty	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
47	month	0	0.1114	0	0.0965	0.1127	0	0.1093	0.0913	0.1305	0	0.1026	0	0.1138	0.1211	0.1162	0.1205	0.0952	0.3996	0.1122	0.126
48	father	0.2197	0.3322	0.1789	0.2915	0.1883	0.2546	0.3465	0.2166	0.2091	0.2619	0.3126	0	0.1896	0.1982	0.3393	0.2926	0.1659	0.323	0.1602	0.3196
49	immunity	0	0.4815	0	0.319	0.0915	0	0.4231	0.2366	0.1029	0	1	0	0.0922	0.0969	0.3305	0.3023	0.0796	0.3571	0.0912	0.1
50	caption	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
51	capital	0.1381	0.098	0.1568	0.0864	0.0991	0.1607	0.3066	0.0822	0.1126	0.1653	0.0912	0	0.0999	0.3147	0.6066	0.4747	0.0853	0.0994	0.5871	0.7312
52	publicity	0	0.0961	0	0.0849	0.0972	0	0.4901	0.0808	0.11	0	0.0895	0	0.0979	0.5058	0.4873	0.1029	0.0839	0.0974	0.0968	0.1068
53	patient	0.2134	0	0.1731	0	0	0.2461	0.1425	0	0	0.2529	0	0	0.0897	0	0.198	0.1445	0	0.0793	0.1555	0
54	arrest	0	0.3394	0	1	0.2895	0	0.3011	0.2206	0.3233	0	0.319	0	0.2524	0.3057	0.307	0.2825	0.2536	0.3298	0.0864	0.2701
55	order	0	0.4562	0	0.3381	0.3209	0	0.8255	0.2505	0.3629	0	0.428	0	0.28	0.7194	0.4012	0.8141	0.3391	0.3812	0.4268	0.5188
56	advice	0	0.0996	0	0.0876	0.1007	0	0.5077	0.0833	0.1147	0	0.0926	0	0.1016	0.5245	0.5047	0.1069	0.0865	0.101	0.1004	0.1112
57	government	0	0.1084	0	0.2701	0.3071	0	0.3636	0.0893	0.3454	0	0.1	0	0.2678	0.3254	0.2875	0.5296	0.267	0.4572	0.7486	1
58	lawmaker	0.1998	0	0.1608	0	0	0.2282	0.134	0	0	0.234	0	0	0.0824	0	0.1862	0.1344	0	0.0736	0.1455	0
59	bed	0.1516	0	0.3106	0	0	0.1746	0.3542	0	0	0.1793	0	0	0.0882	0	0.1408	0.3878	0	0.0782	0.5684	0
60	result	0.0739	0.0992	0.0841	0.2524	0.2901	0.0863	0.5033	0.083	0.33	0.0888	0.0922	0	1	0.5199	0.6149	0.1065	0.2493	0.6311	0.0999	0.2678
61	water	0.1277	0.1054	0.2566	0.092	0.1067	0.1436	0.2184	0.0873	0.1224	0.1468	0.0975	0	0.1076	0.1141	0.4418	0.3657	0.0908	0.1069	0.2349	0.1184
62	cannon	0.1404	0	0.2851	0	0	0.1599	0.3298	0	0	0.1638	0	0	0.08	0	0.1311	0.131	0	0.0716	0.2586	0
63	decade	0	0.0963	0	0.085	0.0973	0	0.0948	0.081	0.1103	0	0.0897	0	0.0981	0.1035	0.0999	0.1031	0.084	0.348	0.097	0.107
64	application	0	0.0965	0	0.2845	0.3258	0	0.4498	0.0811	0.3692	0	0.0898	0	0.2843	0.8895	0.5034	0.1033	0.365	0.2826	0.0971	0.3014
65	regime	0	0.1084	0	0.1624	0.1816	0	0.4004	0.0893	0.2009	0	0.1	0	0.1828	0.1909	0.1715	0.5296	0.1607	0.2474	0.7486	1
66	delegation	0	0.0788	0	0.0711	0.0795	0	0.0778	0.0683	0.088	0	0.0743	0	0.0801	0.0836	0.0812	0.3771	0.0704	0.0797	0.3588	0.4291
67	place	0.1378	0.4683	0.1564	0.3114	0.332	0.1602	0.4129	0.2311	0.3772	0.1649	0.4387	0	0.2898	0.3535	0.7198	0.4335	0.371	0.3476	0.5396	0.3068
68	politician	0.2111	0	0.171	0	0	0.2431	0.141	0	0	0.2497	0	0	0.0884	0	0.196	0.1428	0	0.0783	0.1538	0
69	scene	0.1373	0.3742	0.2781	0.3234	0.2689	0.1559	0.4115	0.2398	0.3028	0.1596	0.3495	0	0.4904	0.3495	0.5523	0.4352	0.2334	0.4878	0.4568	0.3741
70	decision	0	0.1027	0	0.3006	0.5466	0	0.4091	0.0854	0.9715	0	0.0952	0	0.8207	0.3706	0.3223	0.1105	0.2968	0.4324	0.1035	0.3721
71	judge	0.1951	0	0.1566	0	0	0.2222	0.1311	0	0	0.5364	0	0	0.08	0	0.1822	0.131	0	0.0716	0.1421	0
72	evening	0	0.0966	0	0.0853	0.0977	0	0.0951	0.0812	0.1107	0	0.09	0	0.0985	0.1038	0.1002	0.1035	0.0842	0.3491	0.0973	0.1074
73	dictator	1	0	0.1461	0	0	0.2071	0.1237	0	0	0.2118	0	0	0.0739	0	0.1719	0.1224	0	0.0668	0.1334	0
74	iron	0.1277	0.092	0.2566	0.0817	0.093	0.1436	0.3017	0.0779	0.1047	0.1468	0.086	0	0.0937	0.0986	0.3837	0.325	0.0807	0.0932	0.2349	0.1018
75	fist	0.0642	0	0.0717	0	0	0.0732	0.0598	0	0	0.0751	0	0	0.0819	0	0.0649	0.0702	0	0.0731	0.0649	0
76	visitor	0.2062	0	0.1666	0	0	0.2367	0.138	0	0	0.243	0	0	0.0858	0	0.1918	0.1392	0	0.0763	0.1502	0
77	diplomat	0.1928	0	0.1545	0	0	0.2191	0.1296	0	0	0.5288	0	0	0.0787	0	0.1801	0.1293	0	0.0706	0.1403	0
78	policy	0	0.1003	0	0.2012	0.2317	0	0.4855	0.0838	0.2639	0	0.0932	0	0.2336	0.4453	0.5295	0.1077	0.1987	0.305	0.1011	0.3694
79	trial	0	0.0991	0	0.2913	0.3348	0	0.3301	0.0829	0.3808	0	0.0921	0	0.2922	0.3566	0.4047	0.1063	0.3737	0.2905	0.0998	0.4332
80	black	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
81	police	0	0.0982	0	0.0865	0.0992	0	0.0966	0.0823	0.1127	0	0.0913	0	0.1001	0.1056	0.1019	0.4762	0.0854	0.0995	0.4473	0.5463

82	guard	0.2023	0.2426	0.262	0.2588	0.4605	0.2316	0.3894	0.21	0.8644	0.2375	0.2288	0	0.255	0.309	0.8362	0.3916	0.3055	0.2537	0.3719	0.4462
83	stuff	0.1194	0.2426	0.1331	0.2138	0.1263	0.1359	0.4031	0.3593	0.149	0.1393	0.2288	0	0.1276	0.4137	0.5261	0.4216	0.1047	0.2361	0.1257	0.1432
84	idea	0	0.1292	0	0.2504	0.2993	0	0.6459	0.103	0.3555	0	0.1176	0	0.3026	0.3387	0.3278	0.1418	0.2466	0.3769	0.1304	0.4774
85	business	0	0.1128	0	0.3261	0.3816	0	0.3776	0.0923	0.4425	0	0.1038	0	0.3334	0.4101	0.5105	0.5272	0.4177	0.3312	0.492	0.6072
86	army	0	0.1001	0	0.088	0.1013	0	0.0985	0.0837	0.1153	0	0.093	0	0.1021	0.1079	0.104	0.6049	0.0869	0.1015	0.5743	0.5585
87	people	0	0.1161	0	0.1001	0.1176	0	0.1139	0.0945	0.137	0	0.1066	0	0.1188	0.1267	0.1214	0.4648	0.0987	0.118	0.4314	0.4868
88	administration	0	0.0966	0	0.2701	0.3071	0	0.2046	0.0812	0.3454	0	0.09	0	0.2678	0.3254	0.4733	0.4682	0.3214	0.2973	0.4402	1
89	government	0	0.1084	0	0.2701	0.3071	0	0.3636	0.0893	0.3454	0	0.1	0	0.2678	0.3254	0.2875	0.5296	0.267	0.4572	0.7486	1
90	report	0	0.0991	0	0.2861	0.3279	0	0.4825	0.0829	0.3719	0	0.0921	0	0.4345	0.7361	0.4798	0.1063	0.2826	0.4324	0.0998	0.3032
91	disappearance	0	0.0848	0	0.2536	0.7024	0	0.1909	0.0727	0.5186	0	0.0796	0	0.4345	0.3017	0.2689	0.09	0.2509	0.4324	0.0853	0.267
92	passport	0	0.0817	0	0.0734	0.0824	0	0.4176	0.0704	0.0915	0	0.0769	0	0.083	0.4289	0.4155	0.0866	0.0727	0.0826	0.0822	0.0893
93	family	0.1774	0.101	0.1409	0.0887	0.1022	0.1995	0.1199	0.0843	0.1165	0.204	0.0938	0	0.103	0.109	0.1667	0.4903	0.0876	0.1024	0.4597	0.5631
94	total	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
95	police	0	0.0982	0	0.0865	0.0992	0	0.0966	0.0823	0.1127	0	0.0913	0	0.1001	0.1056	0.1019	0.4762	0.0854	0.0995	0.4473	0.5463
96	gas	0.1301	0.4191	0.262	0.2825	0.089	0.1467	0.3741	0.21	0.0997	0.15	0.3952	0	0.5886	0.0941	0.4932	0.3128	0.0777	0.5547	0.2395	0.097
97	allegation	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
98	news	0	0.0976	0	0.2195	0.2475	0	0.4977	0.0819	0.276	0	0.0908	0	0.2493	0.5138	0.4948	0.1046	0.2172	0.248	0.0983	0.2311
99	court	0.1334	0.098	0.5538	0.0864	0.0991	0.1508	0.2275	0.0822	0.1126	0.1543	0.0912	0	0.0999	0.1055	0.125	0.7345	0.0853	0.0994	1	0.7486
100	strongman	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
101	rights	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
102	abuse	0	0.0906	0	0.2691	0.3058	0	0.4176	0.0769	0.3437	0	0.0847	0	0.2667	0.4289	0.4155	0.0966	0.3455	0.2652	0.0912	0.2842
103	country	0.1488	0.1048	0.1707	0.0915	0.106	0.1753	0.1367	0.0869	0.1215	0.1809	0.097	0	0.1069	0.1133	0.1367	0.6103	0.0903	0.1063	0.4827	0.5874
104	demonstrator	0.1719	0	0.1361	0	0	0.1925	0.1164	0	0	0.1967	0	0	0.0682	0	0.1618	0.1141	0	0.0621	0.125	0
105	operation	0.0625	0.3666	0.0696	0.3177	0.3567	0.0711	0.4275	0.2357	0.4094	0.0728	0.3429	0	0.5274	0.3816	0.8589	0.3011	1	0.4729	0.1236	0.3732
106	official	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
107	event	0.0739	0.519	0.0841	0.3633	0.4469	0.0863	0.4518	0.252	0.5492	0.0888	0.4828	0	1	0.4934	0.5492	0.3213	0.3568	0.6311	0.1536	0.3961
108	dissident	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
109	television	0.1434	0	0.2919	0	0	0.1638	0.3363	0	0	0.1679	0	0	0.0821	0	0.1337	0.1341	0	0.0734	0.2642	0
110	interview	0	0.0875	0	0.2608	0.2951	0	0.3334	0.0746	0.3302	0	0.082	0	0.2572	0.8242	0.3317	0.093	0.2579	0.2559	0.088	0.2749

111	division	0	0.1079	0	0.2711	0.544	0	0.6091	0.089	0.6121	0	0.0996	0	0.2689	0.5895	0.2886	0.4992	0.268	0.3243	0.4676	0.5738
112	riot	0	0.0788	0	0.2376	0.4184	0	0.1777	0.0683	0.4628	0	0.0743	0	0.2315	0.2793	0.3259	0.0833	0.3055	0.2305	0.0793	0.2493
113	police	0	0.0982	0	0.0865	0.0992	0	0.0966	0.0823	0.1127	0	0.0913	0	0.1001	0.1056	0.1019	0.4762	0.0854	0.0995	0.4473	0.5463
114	police	0	0.0982	0	0.0865	0.0992	0	0.0966	0.0823	0.1127	0	0.0913	0	0.1001	0.1056	0.1019	0.4762	0.0854	0.0995	0.4473	0.5463
115	line	0.1471	0.323	0.3003	0.3261	0.3816	0.1686	0.4115	0.2677	0.4425	0.173	0.2989	0	0.5734	0.4101	0.4971	0.4629	0.4177	0.5418	0.4355	0.523
116	spokesman	0.1719	0	0.1361	0	0	0.1925	0.1164	0	0	0.7664	0	0	0.0682	0	0.1618	0.1141	0	0.0621	0.125	0
117	response	0.0673	0.0991	0.0757	0.2745	0.3127	0.0774	0.505	0.0829	0.3525	0.0795	0.0921	0	0.7996	0.7	0.6169	0.3071	0.2713	0.5176	0.2922	0.3165
118	death	0.0668	0.3829	0.075	0.3298	0.2884	0.0766	0.3349	0.2445	0.3279	0.0787	0.3571	0	0.6311	0.3071	0.5212	0.312	0.248	1	0.0994	0.4572
119	pressure	0.0695	0.4507	0.0784	0.3011	0.2925	0.0803	0.3991	0.4215	0.327	0.0825	0.4231	0	0.665	0.309	0.5587	0.2862	0.2559	0.6183	0.0957	0.2727
120	release	0.1346	0.0897	0.2719	0.2668	0.7281	0.1524	0.4176	0.0763	0.5397	0.1559	0.0839	0	0.4992	0.4289	0.5106	0.1251	0.2638	0.8263	0.2477	0.2816
121	injury	0	0.5268	0	0.3446	0.2755	0	0.4577	0.2552	0.3059	0	0.4896	0	0.4723	0.2901	0.3581	0.3252	0.3154	0.6244	0.0998	0.2579
122	violation	0	0.0875	0	0.2608	0.2951	0	0.1968	0.0746	0.3302	0	0.082	0	0.2572	0.3119	0.3596	0.093	0.3349	0.2559	0.088	0.2749
123	attitude	0	0.2987	0	0.2562	0.2658	0	0.3906	0.2508	0.2993	0	0.278	0	0.2609	0.2793	0.3259	0.242	0.3055	0.3342	0.1127	0.3448
124	lobby	0.1438	0	0.4628	0	0	0.1642	0.244	0	0	0.1684	0	0	0.0824	0	0.134	0.1344	0	0.0736	0.6168	0
125	group	0.0612	0.1433	0.0681	0.1196	0.1456	0.0695	0.1399	0.1118	0.1766	0.0711	0.1291	0	0.1474	0.1598	0.1514	0.5856	0.1176	0.1461	0.5335	0.6208
126	placard	0	0.0817	0	0.0734	0.0824	0	0.3176	0.0704	0.0915	0	0.0769	0	0.083	0.3262	0.3161	0.0866	0.0727	0.0826	0.0822	0.0893
127	bearing	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
128	condition	0	0.6821	0	0.4279	0.213	0	0.5705	0.3152	0.24	0	0.6209	0	0.2147	0.51	0.6036	0.3984	0.1849	0.4994	0.1295	0.2917
129	son	0.2116	0.0965	0.1715	0.1944	0.2226	0.2438	0.4072	0.0811	0.2523	0.2505	0.0898	0	0.2244	0.2367	0.2076	0.1432	0.1921	0.2949	0.1542	0.3706
130	year	0	0.1143	0	0.0987	0.1158	0	0.1122	0.0933	0.1346	0	0.1051	0	0.1169	0.1246	0.1194	0.6291	0.0974	0.4097	0.5961	0.4208
131	demonstration	0	0.1004	0	0.2947	0.3393	0	0.377	0.0839	0.3866	0	0.0932	0	0.2962	0.3892	0.4097	0.1078	0.378	0.2944	0.1011	0.4572
132	anonymity	0	0.3206	0	0.2825	0.0795	0	0.2862	0.21	0.088	0	0.3023	0	0.0801	0.0836	0.2915	0.2693	0.0704	0.312	0.0793	0.0859
133	reign	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
134	hospital	0.1461	0	1	0	0	0.1673	0.2477	0	0	0.1717	0	0	0.0841	0	0.1361	0.1369	0	0.075	0.5538	0
135	protest	0	0.0923	0	0.2737	0.3117	0	0.2076	0.0781	0.3512	0	0.0862	0	0.2719	0.6801	0.2915	0.0986	0.2705	0.2704	0.093	0.5027

E) HSO measure

No	OC	CC1	CC2	CC3	CC4	CC5	CC6	CC7	CC8	CC9	CC10	CC11	CC12	CC13	CC14	CC15	CC16	CC17	CC18	CC19	CC20
		dictator	confusion	hospital	arrest	murder	citizen	rule	legality	move	supporter	immunity	extradition	outcome	request	charge	community	surgery	death	court	government
1	defense	0	0	3	0	0	0	3	0	3	2	0	0	2	3	5	4	3	0	4	4
2	minister	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
3	priest	2	2	0	2	0	2	0	0	0	2	2	0	0	0	0	0	0	3	0	0
4	magistrate	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
5	murder	0	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0
6	day	2	0	0	0	0	2	4	0	0	2	0	0	0	0	0	2	0	2	0	0
7	embassy	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
8	police	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	3
9	source	3	0	0	0	0	3	3	0	0	3	0	0	3	2	0	2	0	0	2	0
10	insult	0	0	0	0	0	0	3	0	2	0	0	0	0	4	2	0	0	0	0	0
11	team	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	3	3
12	claim	0	0	0	0	0	0	3	0	3	0	0	0	0	6	5	0	2	0	0	0
13	arrest	0	3	0	16	0	0	2	0	0	0	3	0	0	2	0	0	0	4	0	0
14	warrant	0	0	0	0	0	0	3	0	0	0	0	0	0	4	2	0	0	0	0	0
15	attorney	2	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0
16	envoy	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
17	extradition	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	0	0	0	0
18	request	0	2	0	2	0	0	4	0	2	0	0	0	0	16	4	2	2	0	0	2
19	opinion	0	0	0	0	0	0	4	0	0	0	0	0	0	5	3	2	0	0	0	2
20	term	0	0	0	0	0	0	4	0	0	0	0	0	0	4	4	2	0	4	0	0

21	rally	0	2	0	2	0	0	2	0	2	0	0	0	2	3	2	5	2	2	4	4
22	week	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
23	boss	3	0	2	0	0	3	0	0	0	3	0	0	0	0	0	0	2	0	3	0
24	government	0	0	2	0	0	0	0	0	0	0	0	0	0	2	0	3	0	0	4	16
25	official	3	0	0	0	0	3	0	0	0	4	0	0	0	0	0	0	0	0	0	0
26	post	2	2	2	2	0	2	3	0	3	3	0	0	0	3	3	4	2	0	2	2
27	crime	0	0	0	0	0	0	2	0	2	0	0	0	0	0	2	0	0	0	0	0
28	prosecution	0	0	0	0	0	0	3	0	3	2	0	0	0	2	3	4	2	0	0	2
29	institution	0	0	6	0	0	0	4	0	3	0	0	0	0	0	0	3	4	2	4	4
30	power	3	4	2	4	0	3	4	4	0	3	4	0	3	0	4	2	0	5	0	4
31	abuse	0	0	0	0	0	0	3	0	3	0	0	0	0	4	3	0	2	0	0	0
32	constitution	0	2	0	2	0	0	2	2	3	0	3	0	0	3	3	2	0	3	0	0
33	statement	0	2	0	2	0	0	4	0	2	0	0	0	0	5	6	2	2	0	0	2
34	word	0	0	0	0	0	0	3	0	0	0	0	0	0	4	4	2	0	0	0	0
35	military	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
36	support	0	2	3	2	0	0	4	0	4	0	0	0	0	3	4	0	4	0	4	2
37	portrait	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0
38	extradition	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	0	0	0	0
39	law	0	0	2	0	0	0	5	0	0	2	0	0	0	0	0	4	0	0	2	3
40	protester	4	0	0	0	0	4	0	0	0	4	0	0	0	0	2	0	0	2	2	0
41	arrest	0	3	0	16	0	0	2	0	0	0	3	0	0	2	0	0	0	4	0	0
42	opponent	4	0	0	0	0	4	0	0	0	4	0	0	0	0	2	0	0	2	3	0
43	rightist	2	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0
44	politician	3	0	0	0	0	3	0	0	0	3	0	0	0	0	0	0	0	0	0	0
45	time	2	3	0	3	0	2	4	3	0	2	3	0	3	2	0	3	0	4	0	0
46	amnesty	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
47	month	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	2	0	2	0	0
48	father	3	0	0	0	0	3	0	0	0	3	0	0	0	0	0	0	0	2	0	0
49	immunity	0	3	0	3	0	0	4	0	0	0	16	0	0	0	0	0	0	4	0	0
50	caption	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2	0	0	0	0	0
51	capital	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	2	0	0	4	6
52	publicity	0	2	0	2	0	0	4	4	0	0	2	0	0	5	3	2	0	3	0	0
53	patient	3	0	0	0	0	3	0	0	0	3	0	0	0	0	0	0	0	0	0	0
54	arrest	0	3	0	16	0	0	2	0	0	0	3	0	0	2	0	0	0	4	0	0
55	order	0	5	2	4	0	0	4	2	3	0	5	0	0	4	5	5	2	5	2	3
56	advice	0	0	0	0	0	0	3	0	0	0	0	0	0	4	2	0	0	0	0	0
57	government	0	0	2	0	0	0	0	0	0	0	0	0	0	2	0	3	0	0	4	16
58	lawmaker	3	0	0	0	0	3	0	0	0	3	0	0	0	0	0	0	0	0	0	0
59	bed	0	0	3	0	0	0	0	0	0	3	0	0	0	0	0	0	4	0	4	0

60	result	0	0	0	0	0	0	3	0	0	0	0	0	16	4	4	0	0	3	0	0
61	water	0	0	0	0	0	0	0	0	0	2	0	0	2	0	2	0	0	0	2	0
62	cannon	2	0	0	0	0	2	0	0	5	2	0	0	0	0	0	0	0	0	0	0
63	decade	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
64	application	0	2	0	2	0	0	3	0	4	0	0	0	0	4	3	0	2	0	3	2
65	regime	0	0	2	0	0	0	3	0	0	0	0	0	0	0	0	3	0	0	4	16
66	delegation	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	3	4
67	place	0	3	2	3	0	0	4	0	3	0	5	0	0	2	3	3	2	4	3	0
68	politician	3	0	0	0	0	3	0	0	0	3	0	0	0	0	0	0	0	0	0	0
69	scene	0	3	0	3	0	0	2	0	0	0	3	0	3	0	0	3	0	4	2	0
70	decision	0	0	0	0	0	0	0	0	4	0	0	0	4	2	0	0	0	2	2	0
71	judge	3	0	0	0	0	3	0	0	0	3	0	0	0	0	0	0	0	0	3	0
72	evening	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0
73	dictator	16	0	0	0	0	3	0	0	0	3	0	0	0	0	0	0	0	0	0	0
74	iron	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
75	fist	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
76	visitor	3	0	0	0	0	3	0	0	0	3	0	0	0	0	0	0	0	0	0	0
77	diplomat	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
78	policy	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0
79	trial	0	0	0	0	0	0	3	0	3	0	0	0	2	2	3	0	2	2	0	2
80	black	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0
81	police	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	3
82	guard	2	0	3	0	0	2	2	0	5	2	0	0	0	0	5	0	0	0	5	2
83	stuff	0	2	0	2	0	0	3	4	0	0	2	0	2	4	5	2	0	3	0	0
84	idea	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0
85	business	3	2	2	2	0	3	4	0	4	3	0	0	0	3	4	4	3	0	4	5
86	army	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	4	0	0	3	3
87	people	4	0	0	0	0	5	0	0	0	4	0	0	0	2	2	5	0	2	2	3
88	administration	0	0	2	0	0	0	2	0	0	0	0	0	0	2	0	3	0	0	6	16
89	government	0	0	2	0	0	0	0	0	0	0	0	0	0	2	0	3	0	0	4	16
90	report	0	0	0	0	0	0	3	0	0	0	0	0	2	4	3	0	0	2	0	0
91	disappearance	0	2	0	2	3	0	2	0	3	0	3	0	4	3	2	0	2	4	0	2
92	passport	0	0	0	0	0	0	2	3	0	0	0	0	0	3	0	0	0	2	0	0
93	family	3	0	4	0	0	3	0	0	0	3	0	0	0	0	0	4	0	0	2	3

94	total	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
95	police	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	3
96	gas	0	2	0	2	0	0	3	0	0	0	4	0	2	0	0	0	0	3	0	0
97	allegation	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
98	news	0	0	0	0	0	0	3	2	0	0	0	0	0	4	2	0	0	0	0	0
99	court	0	0	5	0	0	0	0	0	2	0	0	0	0	0	0	4	5	0	16	4
100	strongman	4	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0
101	rights	0	0	0	0	0	0	3	0	3	0	0	0	0	4	3	0	2	0	0	0
102	abuse	3	0	3	0	0	5	0	0	0	3	0	0	0	0	0	4	0	0	4	2
103	country	2	0	0	0	0	2	0	0	0	3	0	0	0	0	0	0	0	0	0	0
104	demonstrator	0	3	0	3	0	0	4	0	4	0	3	0	4	3	6	2	16	4	0	2
105	operation	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0
106	official	0	4	0	4	0	0	4	0	4	0	4	0	16	5	4	3	4	5	3	5
107	event	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
108	dissident	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
109	television	0	0	0	0	0	0	0	0	0	0	0	0	0	5	3	0	0	0	0	0
110	interview	0	3	3	0	0	0	5	0	3	0	0	0	0	3	2	4	0	0	6	5
111	division	0	5	0	3	0	0	2	0	2	0	3	0	0	3	0	1	0	4	0	0
112	riot	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	3
113	police	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	2	3
114	police	0	2	2	2	0	0	4	2	4	3	2	0	0	3	4	3	3	3	3	3
115	line	2	0	0	0	0	2	0	0	0	4	0	0	0	0	0	0	0	0	0	0
116	spokesman	0	2	0	2	0	0	3	0	2	0	2	0	4	5	4	0	2	2	0	2
117	response	0	4	0	4	4	0	3	2	2	0	4	0	3	0	2	2	0	16	0	0
118	death	0	0	0	0	0	0	0	0	0	0	2	0	3	2	4	0	0	3	0	0
119	pressure	0	2	0	2	3	0	4	0	6	0	4	0	3	3	4	0	3	5	2	2
120	release	0	0	0	0	0	0	3	0	3	0	2	0	0	2	3	0	2	2	0	0
121	injury	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
122	violation	0	0	0	0	0	0	3	0	3	0	0	0	0	2	3	0	2	0	0	0
123	attitude	0	0	0	0	0	0	4	0	0	0	2	0	0	0	0	2	0	2	0	0
124	lobby	3	0	5	0	0	3	0	0	0	3	0	0	0	0	0	4	5	0	5	2
125	group	0	2	3	2	0	0	2	2	0	3	2	0	0	3	2	5	0	3	4	5
126	placard	0	0	0	0	0	0	2	0	0	0	0	0	0	3	0	2	0	0	0	0
127	bearing	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
128	condition	0	5	0	4	0	0	6	2	3	0	4	0	0	4	4	2	4	5	0	0
129	son	2	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0	0
130	year	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	5	0	0	4	3
131	demonstration	0	0	0	0	0	0	2	0	0	0	0	0	0	3	0	2	0	0	0	2
132	anonymity	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
133	reign	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
134	hospital	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	5	2
135	protest	0	0	0	0	0	0	0	0	0	0	0	0	0	4	3	0	0	0	0	3

F) LESK measure

No	OC	CC1	CC2	CC3	CC4	CC5	CC6	CC7	CC8	CC9	CC10	CC11	CC12	CC13	CC14	CC15	CC16	CC17	CC18	CC19	CC20
		dictator	confusion	hospital	arrest	murder	citizen	rule	legality	move	supporter	immunity	extradition	outcome	request	charge	community	surgery	death	court	government
1	defense	47	72	68	40	75	73	150	48	133	138	57	41	135	52	78	93	186	66	247	195
2	minister	56	48	56	21	60	95	64	22	58	120	34	29	52	43	62	58	78	43	93	109
3	priest	69	53	75	24	93	113	142	33	118	173	53	44	121	75	88	121	165	80	195	179
4	magistrate	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
5	murder	54	57	55	29	1324	71	86	31	106	105	39	39	85	56	62	73	136	58	102	115
6	day	54	48	65	20	81	79	186	40	174	145	58	27	177	67	96	103	262	71	240	175
7	embassy	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
8	police	43	48	56	24	56	70	85	33	69	95	40	31	69	51	66	96	90	54	123	127
9	source	46	40	50	23	50	78	67	26	72	102	34	29	70	46	63	57	87	58	70	90
10	insult	20	44	29	34	38	41	39	28	51	38	24	23	38	37	44	29	54	37	45	45
11	team	52	56	74	19	93	96	141	25	127	167	67	38	147	70	96	131	213	68	163	175
12	claim	22	39	31	23	40	49	52	37	47	45	30	26	41	31	196	39	56	38	57	64
13	arrest	10	35	14	411	29	23	40	14	49	21	25	18	25	24	29	20	36	39	21	44
14	warrant	48	45	56	15	73	71	79	30	76	100	37	28	76	50	61	74	112	62	135	100
15	attorney	87	60	71	28	99	104	125	45	126	159	56	41	126	70	87	119	168	86	225	184
16	envoy	29	26	22	12	31	49	32	13	24	64	15	18	17	19	35	24	27	19	27	34
17	extradition	17	32	25	18	39	52	36	21	37	45	21	650	38	27	35	31	53	39	41	48
18	request	31	36	48	24	56	53	64	31	73	81	35	27	67	1494	75	48	104	50	88	79
19	opinion	43	54	47	22	58	58	84	32	80	73	52	31	97	130	69	57	100	61	95	108
20	term	50	62	68	22	90	88	155	31	129	123	49	41	136	59	89	104	173	79	151	157
21	rally	21	23	23	15	26	32	42	16	36	29	17	16	21	20	27	110	41	18	53	48
22	week	32	45	49	19	51	55	79	25	58	62	38	30	56	43	55	56	80	51	88	75
23	boss	63	26	30	11	32	68	37	13	29	95	16	21	29	26	39	31	42	25	45	46
24	government	81	71	86	44	115	121	171	39	152	170	70	48	163	79	115	152	204	92	342	3878
25	official	78	48	77	19	82	125	162	28	181	254	60	35	202	60	110	111	293	75	247	207
26	post	55	61	65	41	94	84	137	28	184	144	54	38	160	67	96	95	279	71	155	154
27	crime	59	56	59	48	118	83	113	44	129	133	51	41	116	68	83	88	179	69	152	147
28	prosecution	42	58	67	26	88	80	72	39	70	87	37	47	68	55	77	73	93	59	112	97
29	institution	27	51	137	64	46	55	67	22	77	90	33	27	64	41	60	76	93	57	113	104
30	power	53	64	52	26	91	72	112	54	145	130	43	36	125	68	78	76	163	137	131	138
31	abuse	28	28	38	28	40	38	53	22	54	47	22	20	37	36	40	37	69	41	51	52
32	constitution	42	47	50	63	87	64	100	25	116	90	43	31	102	50	60	73	129	59	153	137

33	statement	52	62	69	25	80	86	165	34	170	161	77	45	208	150	409	100	298	85	191	176
34	word	60	64	76	30	102	99	237	42	233	204	89	42	247	96	126	151	393	108	238	192
35	military	20	29	29	12	24	38	37	21	34	45	19	17	45	27	32	27	61	29	51	42
36	support	46	52	75	43	81	92	177	33	224	238	54	43	201	68	117	93	345	77	186	155
37	portrait	28	43	34	15	48	57	45	18	42	55	21	29	36	30	63	38	54	31	48	52
38	extradition	17	32	25	18	39	52	36	21	37	45	21	650	38	27	35	31	53	39	41	48
39	law	72	77	85	25	106	104	788	43	273	214	91	38	295	88	147	136	520	102	266	233
40	protester	34	26	27	12	45	58	37	13	40	72	22	15	33	27	32	38	47	23	49	47
41	arrest	10	35	14	411	29	23	40	14	49	21	25	18	25	24	29	20	36	39	21	44
42	opponent	43	24	29	9	49	52	36	15	32	76	26	17	38	21	28	33	57	21	47	55
43	rightist	38	36	33	11	41	73	46	15	40	88	25	22	40	26	41	40	55	28	49	70
44	politician	102	52	73	16	86	123	170	23	174	219	48	28	205	63	98	108	276	63	243	188
45	time	42	46	40	23	62	60	111	41	106	103	68	27	113	52	71	78	149	59	138	125
46	amnesty	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
47	month	40	49	48	23	54	64	87	30	61	72	41	35	61	39	64	62	104	59	79	91
48	father	66	38	53	24	69	73	126	28	120	131	41	27	129	51	65	110	167	57	204	165
49	immunity	31	34	38	25	39	50	58	29	50	61	2581	21	70	35	45	42	108	41	71	70
50	caption	32	39	44	17	46	46	70	20	56	55	32	27	53	37	42	54	71	38	62	71
51	capital	41	59	58	24	67	72	75	35	82	75	46	36	76	51	60	65	99	54	99	107
52	publicity	42	61	53	21	53	72	71	31	61	83	39	33	82	120	70	60	88	59	74	83
53	patient	68	57	70	20	88	141	87	17	83	171	52	35	82	52	68	83	125	60	107	123
54	arrest	10	35	14	411	29	23	40	14	49	21	25	18	25	24	29	20	36	39	21	44
55	order	54	62	77	52	86	94	180	37	131	139	63	39	126	136	105	150	179	138	179	154
56	advice	28	29	28	11	24	39	50	14	42	44	23	15	41	34	43	31	60	32	40	46
57	government	81	71	86	44	115	121	171	39	152	170	70	48	163	79	115	152	204	92	342	3878
58	lawmaker	60	28	32	9	32	59	39	18	31	78	22	18	25	21	41	31	45	26	48	45
59	bed	62	77	104	27	97	102	133	36	133	156	68	47	140	81	120	105	219	80	155	158
60	result	54	54	64	26	85	88	199	28	195	173	70	38	11245	67	104	105	256	83	186	163
61	water	87	86	91	26	106	111	183	37	202	166	95	47	212	72	119	123	313	100	206	188
62	cannon	25	35	37	28	44	54	71	26	59	66	50	33	60	42	45	53	116	58	72	77

63	decade	20	32	21	15	36	34	62	21	55	48	26	20	48	29	31	38	82	29	61	57
64	application	40	46	60	48	89	65	111	26	156	110	55	31	111	347	74	72	195	64	119	130
65	regime	81	71	86	35	115	121	171	39	152	170	70	48	163	79	115	152	204	92	342	3878
66	delegation	18	21	17	24	27	26	24	20	39	30	19	14	17	27	28	64	40	23	39	45
67	place	58	70	68	41	101	95	162	42	211	183	65	41	197	84	123	105	279	87	173	173
68	politician	102	52	73	16	86	123	170	23	174	219	48	28	205	63	98	108	276	63	243	188
69	scene	48	64	68	24	70	78	88	29	89	86	51	38	80	50	73	74	119	63	92	109
70	decision	27	46	37	61	76	49	51	29	184	60	32	27	143	56	58	44	106	59	64	78
71	judge	72	47	59	14	78	102	81	27	80	134	45	29	86	48	72	91	124	54	154	126
72	evening	22	29	26	23	37	39	56	20	48	50	34	26	48	34	42	51	67	41	55	68
73	dictator	2595	35	44	10	54	66	62	26	48	96	31	17	54	31	40	50	73	32	70	81
74	iron	55	63	75	26	68	79	114	30	89	89	75	31	110	57	74	79	156	69	135	134
75	fist	10	13	15	10	14	18	21	8	29	21	11	10	20	17	19	20	31	16	24	29
76	visitor	31	19	26	6	28	48	27	10	20	61	15	15	22	18	27	26	35	17	36	40
77	diplomat	96	42	53	16	74	83	144	23	137	158	39	34	151	45	71	103	224	62	230	179
78	policy	43	48	66	24	73	71	94	26	89	92	38	37	83	53	65	83	117	62	116	127
79	trial	74	90	84	33	120	107	138	58	123	129	74	58	128	82	98	95	177	89	176	165
80	black	24	37	31	15	34	44	49	86	43	55	25	22	38	32	35	51	67	36	55	61
81	police	43	48	56	24	56	70	85	33	69	95	40	31	69	51	66	96	90	54	123	127
82	guard	57	37	44	16	48	95	49	18	42	122	30	22	46	34	181	49	69	33	60	79
83	stuff	62	73	75	29	104	93	177	38	184	172	74	45	194	77	119	119	337	87	201	173
84	idea	37	51	45	24	54	60	182	32	89	97	48	28	109	60	84	65	128	60	99	108
85	business	53	60	76	28	88	76	104	34	105	108	51	37	105	66	100	92	136	61	135	172
86	army	49	53	70	24	56	65	92	27	86	121	48	36	78	52	67	81	114	59	196	170
87	people	68	56	71	28	108	321	169	34	158	162	54	52	178	77	97	186	228	67	200	209
88	administration	59	66	74	48	82	99	111	46	96	114	53	59	95	68	79	120	128	75	196	719
89	government	81	71	86	44	115	121	171	39	152	170	70	48	163	79	115	152	204	92	342	3878
90	report	42	44	49	45	79	61	91	28	84	88	49	34	93	61	73	76	135	64	102	126
91	disappearance	15	29	26	38	33	36	41	16	46	38	23	22	41	28	29	29	45	55	43	47
92	passport	24	29	29	15	32	47	42	19	36	41	29	21	38	27	40	33	52	27	52	56
93	family	60	51	67	17	85	79	123	29	124	117	46	33	132	49	79	115	178	72	195	135

94	total	8	11	10	9	12	15	20	12	14	16	9	12	15	10	17	11	22	15	18	18
95	police	43	48	56	24	56	70	85	33	69	95	40	31	69	51	66	96	90	54	123	127
96	gas	62	67	77	30	93	102	195	36	191	181	78	34	224	77	120	120	338	85	225	184
97	allegation	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
98	news	28	36	35	24	33	42	57	30	53	53	29	20	56	28	53	45	80	41	59	70
99	court	70	60	109	21	102	120	186	47	164	197	69	41	186	88	115	156	250	85	7009	342
100	strongman	233	19	18	11	20	29	26	22	19	31	15	13	23	16	20	18	28	17	27	34
101	rights	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
102	abuse	28	28	38	28	40	38	53	22	54	47	22	20	37	36	40	37	69	41	51	52
103	country	83	82	81	33	108	116	196	37	246	196	75	50	254	76	113	145	434	89	267	218
104	demonstrator	34	23	27	10	32	33	31	13	28	47	21	15	24	20	28	38	40	15	48	40
105	operation	73	74	103	80	136	115	236	41	346	212	108	53	256	104	149	144	18283	107	250	204
106	official	20	19	21	8	24	32	27	33	28	46	23	17	29	24	24	19	48	19	40	40
107	event	54	55	64	25	85	88	199	28	195	173	70	39	11245	67	104	105	256	88	186	163
108	dissident	9	17	16	5	20	22	21	9	15	19	7	13	14	10	26	14	28	13	21	21
109	television	52	59	73	20	66	87	124	34	84	97	73	33	99	59	88	82	164	69	129	146
110	interview	24	38	31	15	35	46	38	20	39	46	25	23	31	26	39	32	47	31	43	45
111	division	46	52	72	61	86	65	168	30	120	107	45	39	128	59	75	123	193	77	148	245
112	riot	31	106	27	26	46	52	30	18	41	59	29	23	37	32	41	40	51	26	45	41
113	police	43	48	56	24	56	70	85	33	69	95	40	31	69	51	66	96	90	54	123	127
114	police	43	48	56	24	56	70	85	33	69	95	40	31	69	51	66	96	90	54	123	127
115	line	62	69	87	23	104	104	187	36	205	162	68	46	188	79	126	114	277	99	190	182
116	spokesman	8	5	6	2	6	10	7	3	6	15	7	4	5	6	7	6	11	7	11	11
117	response	50	56	55	30	78	87	130	36	144	143	82	33	150	108	92	89	205	82	161	145
118	death	32	46	47	39	58	59	72	26	71	75	41	39	83	50	69	59	107	1987	85	92
119	pressure	49	56	55	64	93	67	163	32	157	127	53	34	160	62	89	96	198	69	156	148
120	release	45	45	45	45	92	56	83	30	130	96	62	27	86	62	56	61	149	90	105	99
121	injury	57	68	58	29	103	88	137	32	170	150	67	33	182	67	98	86	295	71	170	157
122	violation	39	46	42	22	48	66	62	28	57	89	40	25	64	39	49	51	93	46	83	78
123	attitude	49	63	61	24	75	90	145	30	186	128	44	33	139	63	97	98	214	64	157	143
124	lobby	27	35	37	12	40	52	48	18	36	51	25	23	43	24	36	47	162	27	351	56
125	group	60	62	73	22	89	161	152	36	175	137	56	41	193	66	99	341	268	77	247	202

126	placard	26	33	38	13	42	45	49	19	44	51	33	22	42	33	41	39	63	34	55	59
127	bearing	25	25	33	14	27	38	52	19	39	47	21	15	44	25	33	51	70	28	58	46
128	condition	64	83	75	49	94	111	214	40	260	200	205	42	250	68	121	126	476	169	216	201
129	son	34	29	35	18	40	41	52	17	43	52	27	20	49	33	35	46	60	37	63	70
130	year	42	53	51	25	55	64	89	33	87	81	45	36	77	51	74	121	110	62	103	109
131	demonstration	45	53	43	64	73	62	60	32	111	70	40	36	58	56	60	50	109	59	69	80
132	anonymity	4	19	5	27	9	10	22	5	11	9	13	6	9	8	9	8	10	15	8	15
133	reign	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
134	hospital	44	46	1614	14	55	65	64	20	56	81	38	25	64	48	64	70	159	47	109	86
135	protest	31	37	40	41	67	47	57	21	73	62	27	24	52	48	51	48	83	38	64	71

4. Results for Ideal Summaries with diffrents cutting points

" for full data, please contact the author email[9]"

a) WuP Results First Data

					WuP with different cutting point									
File Name	Text	OT size	CT size	Human Ev.%	40% cutting point	50% cutting point	60% cutting point	70% cutting point	80% cutting point	ERROR 40%	ERROR 50%	ERROR 60%	ERROR 70%	ERROR 80%
APW19981019.0098	T1	8 KB	1 KB	100	90.37	90.37	90.37	87.41	64.44	**93**	**93**	**93**	**159**	**1264**
APW19981022.0269	T2	4KB	1 KB	100	87.36	87.36	86.21	80.46	60.92	**160**	**160**	**190**	**382**	**1527**
APW19981030.0792	T3	5KB	1KB	100	86.26	86.26	83.97	79.39	62.60	**189**	**189**	**257**	**425**	**1399**
APW19981120.0290	T4	5KB	1KB	100	87.88	87.12	84.85	79.55	66.67	**147**	**166**	**230**	**418**	**1111**
APW19981202.1274	T5	5KB	1KB	100	88.12	88.12	88.12	86.14	61.39	**141**	**141**	**141**	**192**	**1491**
APW19981211.1276	T6	6KB	1KB	100	89.29	89.29	89.29	78.57	52.68	**115**	**115**	**115**	**459**	**2239**
APW19981212.0161	T7	6KB	1KB	100	85.33	85.33	84.78	79.35	60.33	**215**	**215**	**232**	**427**	**1574**
APW19981227.0870	T8	4KB	1KB	100	86.42	86.42	85.19	76.54	59.26	**184**	**184**	**219**	**550**	**1660**
NYT19981001.0379	T9	20KB	1KB	100	89.67	88.66	85.14	80.10	63.73	**107**	**128**	**221**	**396**	**1316**
NYT19981003.0120	T10	6KB	1KB	100	91.14	91.14	86.71	80.38	56.96	**79**	**79**	**177**	**385**	**1852**
NYT19981004.0102	T11	9KB	1KB	100	84.97	84.97	84.02	76.68	66.84	**226**	**226**	**255**	**544**	**1100**
NYT19981010.0149	T12	9KB	1KB	100	85.81	85.81	85.14	85.14	74.32	**201**	**201**	**221**	**221**	**659**
NYT19981012.0334	T13	10KB	1KB	100	84.62	83.26	76.47	71.49	48.87	**237**	**280**	**554**	**813**	**2614**
NYT19981012.0359	T14	9KB	1KB	100	85.19	84.13	78.84	75.13	55.56	**219**	**252**	**448**	**618**	**1975**
NYT19981013.0354	T15	8KB	1KB	100	92.61	92.05	90.91	84.09	75.00	**55**	**63**	**83**	**253**	**625**
NYT19981013.0399	T16	8KB	1KB	100	92.55	91.93	88.20	85.09	61.49	**56**	**65**	**139**	**222**	**1483**
NYT19981017.0027	T17	9KB	1KB	100	80.31	80.31	78.74	75.59	61.42	**388**	**388**	**452**	**596**	**1489**
NYT19981018.0091	T18	10KB	1KB	100	91.89	90.81	87.03	85.41	76.22	**66**	**84**	**168**	**213**	**566**
NYT19981024.0050	T19	11KB	1KB	100	92.79	92.79	84.62	78.85	59.62	**52**	**52**	**237**	**447**	**1631**
NYT19981104.0545	T20	11KB	1KB	100	82.55	81.13	78.77	72.64	51.42	**305**	**356**	**451**	**748**	**2360**
NYT19981105.0538	T21	8KB	1KB	100	90.72	90.21	86.60	78.87	58.25	**86**	**96**	**180**	**447**	**1743**
NYT19981107.0056	T22	7Kb	1KB	100	85.88	85.29	84.12	78.82	54.71	**199**	**216**	**252**	**448**	**2052**
NYT19981107.0251	T23	5Kb	1KB	100	86.02	86.02	86.02	79.57	66.67	**195**	**195**	**195**	**417**	**1111**
NYT19981114.0079	T24	13KB	1KB	100	88.43	88.43	85.45	78.36	57.46	**134**	**134**	**212**	**468**	**1809**
NYT19981114.0129	T25	15KB	1KB	100	92.43	91.35	89.73	85.95	71.89	**57**	**75**	**105**	**198**	**790**
NYT19981121.0117	T26	7KB	1KB	100	89.40	89.40	86.09	70.20	56.95	**112**	**112**	**193**	**888**	**1853**
NYT19981122.0163	T27	9KB	1KB	100	90.72	90.72	90.21	87.11	79.38	**86**	**86**	**96**	**166**	**425**
NYT19981126.0192	T28	9KB	1Kb	100	92.43	91.35	89.73	85.95	71.89	**57**	**75**	**105**	**198**	**790**
NYT19981201.0444	T29	7KB	1KB	100	84.52	82.58	72.26	68.39	47.74	**240**	**303**	**770**	**999**	**2731**
NYT19981204.0365	T30	7KB	1KB	100	89.50	88.00	87.00	76.00	66.00	**110**	**144**	**169**	**576**	**1156**
NYT19981209.0451	T31	13KB	1KB	100	91.04	90.67	90.30	86.57	73.13	**80**	**87**	**94**	**180**	**722**
NYT19981219.0117	T32	13KB	1KB	100	82.13	81.70	77.45	71.49	44.26	**319**	**335**	**509**	**813**	**3107**
NYT19981221.0377	T33	8KB	1KB	100	88.76	88.76	88.76	84.83	69.10	**126**	**126**	**126**	**230**	**955**
NYT19981223.0347	T34	11KB	1KB	100	93.89	93.89	93.01	88.21	68.56	**37**	**37**	**49**	**139**	**989**
										149	**161**	**233**	**430**	**1476**
										12.22	12.67	15.28	20.75	38.41

[9] Zainab_bayram@hotmail.com

b) LCH Results First Data

					LCH with different cutting point									
File Name	Text	OT size	CT size	Human Ev. %	40% cutting point	50% cutting point	60% cutting point	70% cutting point	80% cutting point	ERROR 40%	ERROR 50%	ERROR 60%	ERROR 70%	ERROR 80%
APW19981019.0098	T1	8 KB	1 KB	100	91.11	91.11	89.63	88.89	81.48	**79**	**79**	**108**	**123**	**343**
APW19981022.0269	T2	4KB	1 KB	100	87.36	87.36	86.21	80.46	71.26	**160**	**160**	**190**	**382**	**826**
APW19981030.0792	T3	5KB	1KB	100	86.26	86.26	86.26	82.44	74.05	**189**	**189**	**189**	**308**	**674**
APW19981120.0290	T4	5KB	1KB	100	87.88	87.88	87.88	84.09	62.88	**147**	**147**	**147**	**253**	**1378**
APW19981202.1274	T5	5KB	1KB	100	88.12	88.12	88.12	88.12	78.22	**141**	**141**	**141**	**141**	**474**
APW19981211.1276	T6	6KB	1KB	100	89.29	89.29	88.39	85.71	64.29	**115**	**115**	**135**	**204**	**1276**
APW19981212.0161	T7	6KB	1KB	100	88.59	88.59	88.04	84.24	78.26	**130**	**130**	**143**	**248**	**473**
APW19981227.0870	T8	4KB	1KB	100	86.42	86.42	85.19	81.48	70.37	**184**	**184**	**219**	**343**	**878**
NYT19981001.0379	T9	20KB	1KB	100	90.18	89.92	89.17	87.66	78.59	**97**	**102**	**117**	**152**	**458**
NYT19981003.0120	T10	6KB	1KB	100	91.77	91.77	91.77	89.87	83.54	**68**	**68**	**68**	**103**	**271**
NYT19981004.0102	T11	9KB	1KB	100	84.97	84.97	84.02	81.87	74.09	**226**	**226**	**255**	**329**	**671**
NYT19981010.0149	T12	9KB	1KB	100	85.81	85.81	85.14	84.46	71.62	**201**	**201**	**221**	**242**	**805**
NYT19981012.0334	T13	10KB	1KB	100	85.97	85.97	84.16	75.11	73.30	**197**	**197**	**251**	**619**	**713**
NYT19981012.0359	T14	9KB	1KB	100	86.24	86.24	84.13	80.42	69.31	**189**	**189**	**252**	**383**	**942**
NYT19981013.0354	T15	8KB	1KB	100	93.18	93.18	92.05	86.93	73.86	**46**	**46**	**63**	**171**	**683**
NYT19981013.0399	T16	8KB	1KB	100	92.55	92.55	92.55	91.30	79.50	**56**	**56**	**56**	**76**	**420**
NYT19981017.0027	T17	9KB	1KB	100	80.31	80.31	80.31	77.17	72.44	**388**	**388**	**388**	**521**	**760**
NYT19981018.0091	T18	10KB	1KB	100	92.97	92.97	91.89	86.49	76.22	**49**	**49**	**66**	**183**	**566**
NYT19981024.0050	T19	11KB	1KB	100	93.27	93.27	93.27	89.42	77.40	**45**	**45**	**45**	**112**	**511**
NYT19981104.0545	T20	11KB	1KB	100	82.55	82.55	82.55	79.25	69.81	**305**	**305**	**305**	**431**	**911**
NYT19981105.0538	T21	8KB	1KB	100	90.72	90.72	90.72	88.14	76.80	**86**	**86**	**86**	**141**	**538**
NYT19981107.0056	T22	7Kb	1KB	100	87.06	87.06	86.47	83.53	70.59	**167**	**167**	**183**	**271**	**865**
NYT19981107.0251	T23	5Kb	1KB	100	86.02	86.02	86.02	81.72	73.12	**195**	**195**	**195**	**334**	**723**
NYT19981114.0079	T24	13KB	1KB	100	88.43	88.43	88.43	85.07	67.16	**134**	**134**	**134**	**223**	**1078**
NYT19981114.0129	T25	15KB	1KB	100	92.43	91.89	89.19	86.49	84.86	**57**	**66**	**117**	**183**	**229**
NYT19981121.0117	T26	7KB	1KB	100	89.40	89.40	89.40	86.09	68.87	**112**	**112**	**112**	**193**	**969**
NYT19981122.0163	T27	9KB	1KB	100	90.72	90.72	90.21	83.51	75.26	**86**	**86**	**96**	**272**	**612**
NYT19981126.0192	T28	9KB	1Kb	100	92.43	91.89	89.19	86.49	84.86	**57**	**66**	**117**	**183**	**229**
NYT19981201.0444	T29	7KB	1KB	100	87.10	87.10	87.10	79.35	71.61	**166**	**166**	**166**	**426**	**806**
NYT19981204.0365	T30	7KB	1KB	100	90.00	90.00	89.00	86.50	74.00	**100**	**100**	**121**	**182**	**676**
NYT19981209.0451	T31	13KB	1KB	100	91.04	91.04	89.93	88.06	70.90	**80**	**80**	**101**	**143**	**847**
NYT19981219.0117	T32	13KB	1KB	100	82.13	82.13	82.13	80.85	72.34	**319**	**319**	**319**	**367**	**765**
NYT19981221.0377	T33	8KB	1KB	100	88.76	88.76	87.64	86.52	79.78	**126**	**126**	**153**	**182**	**409**
NYT19981223.0347	T34	11KB	1KB	100	93.89	93.89	93.89	92.58	83.41	**37**	**37**	**37**	**55**	**275**
										139	**140**	**156**	**249**	**678**
										11.80	11.83	12.48	15.79	26.04

c) Resnik Results First Data

					Resnik with different cutting point									
File Name	**Text**	**OT size**	**CT size**	**Human Ev.**	**40% cutting point**	**50% cutting point**	**60% cutting point**	**70% cutting point**	**80% cutting point**	**ERROR 40%**	**ERROR 50%**	**ERROR 60%**	**ERROR 70%**	**ERROR 80%**
APW19981019.0098	T1	8 KB	1 KB	100	85.19	85.19	81.48	68.15	58.52	**219**	**219**	**343**	**1015**	**1721**
APW19981022.0269	T2	4KB	1 KB	100	85.06	82.76	79.31	71.26	60.92	**223**	**297**	**428**	**826**	**1527**
APW19981030.0792	T3	5KB	1KB	100	80.92	74.05	70.23	59.54	55.73	**364**	**674**	**886**	**1637**	**1960**
APW19981120.0290	T4	5KB	1KB	100	82.58	80.30	75.00	68.18	61.36	**304**	**388**	**625**	**1012**	**1493**
APW19981202.1274	T5	5KB	1KB	100	81.19	79.21	74.26	68.32	60.40	**354**	**432**	**663**	**1004**	**1568**
APW19981211.1276	T6	6KB	1KB	100	82.14	77.68	76.79	69.64	63.39	**319**	**498**	**539**	**922**	**1340**
APW19981212.0161	T7	6KB	1KB	100	79.35	75.00	73.91	65.76	58.15	**427**	**625**	**681**	**1172**	**1751**
APW19981227.0870	T8	4KB	1KB	100	83.95	79.01	72.84	67.90	51.85	**258**	**440**	**738**	**1030**	**2318**
NYT19981001.0379	T9	20KB	1KB	100	83.88	81.61	77.33	68.51	61.21	**260**	**338**	**514**	**991**	**1505**
NYT19981003.0120	T10	6KB	1KB	100	87.34	85.44	81.65	72.15	58.23	**160**	**212**	**337**	**776**	**1745**
NYT19981004.0102	T11	9KB	1KB	100	79.79	77.72	73.20	67.36	60.10	**408**	**496**	**718**	**1066**	**1592**
NYT19981010.0149	T12	9KB	1KB	100	78.38	77.03	73.65	67.57	65.54	**467**	**528**	**694**	**1052**	**1187**
NYT19981012.0334	T13	10KB	1KB	100	76.92	76.02	73.30	66.06	55.66	**533**	**575**	**713**	**1152**	**1966**
NYT19981012.0359	T14	9KB	1KB	100	80.42	78.84	74.07	70.37	58.73	**383**	**448**	**672**	**878**	**1703**
NYT19981013.0354	T15	8KB	1KB	100	86.36	85.23	70.45	70.45	64.20	**186**	**218**	**873**	**873**	**1281**
NYT19981013.0399	T16	8KB	1KB	100	87.58	86.34	83.23	73.29	57.76	**154**	**187**	**281**	**713**	**1784**
NYT19981017.0027	T17	9KB	1KB	100	75.59	74.02	69.29	64.57	54.33	**596**	**675**	**943**	**1256**	**2086**
NYT19981018.0091	T18	10KB	1KB	100	87.03	85.41	80.54	71.89	67.03	**168**	**213**	**379**	**790**	**1087**
NYT19981024.0050	T19	11KB	1KB	100	87.02	86.54	79.33	76.44	65.38	**169**	**181**	**427**	**555**	**1198**
NYT19981104.0545	T20	11KB	1KB	100	75.47	74.06	69.34	65.57	50.94	**602**	**673**	**940**	**1186**	**2407**
NYT19981105.0538	T21	8KB	1KB	100	84.02	82.99	76.29	66.49	53.61	**255**	**289**	**562**	**1123**	**2152**
NYT19981107.0056	T22	7Kb	1KB	100	77.65	75.88	71.18	68.24	58.24	**500**	**582**	**831**	**1009**	**1744**
NYT19981107.0251	T23	5Kb	1KB	100	79.57	77.42	72.04	63.44	55.91	**417**	**510**	**782**	**1337**	**1944**
NYT19981114.0079	T24	13KB	1KB	100	84.33	82.46	76.87	70.52	60.45	**246**	**308**	**535**	**869**	**1564**
NYT19981114.0129	T25	15KB	1KB	100	87.57	84.32	77.30	71.35	63.24	**155**	**246**	**515**	**821**	**1351**
NYT19981121.0117	T26	7KB	1KB	100	80.13	78.15	74.17	70.20	62.91	**395**	**478**	**667**	**888**	**1375**
NYT19981122.0163	T27	9KB	1KB	100	85.57	82.47	79.90	72.16	67.01	**208**	**307**	**404**	**775**	**1088**
NYT19981126.0192	T28	9KB	1Kb	100	87.57	84.32	77.30	71.35	63.24	**155**	**246**	**515**	**821**	**1351**
NYT19981201.0444	T29	7KB	1KB	100	74.19	72.90	65.16	60.00	51.61	**666**	**734**	**1214**	**1600**	**2341**
NYT19981204.0365	T30	7KB	1KB	100	81.50	78.50	72.00	67.50	61.00	**342**	**462**	**784**	**1056**	**1521**
NYT19981209.0451	T31	13KB	1KB	100	84.33	83.58	77.61	71.27	65.30	**246**	**270**	**501**	**825**	**1204**
NYT19981219.0117	T32	13KB	1KB	100	74.04	72.77	67.66	63.83	59.57	**674**	**742**	**1046**	**1308**	**1634**
NYT19981221.0377	T33	8KB	1KB	100	78.65	77.53	71.91	66.85	62.36	**456**	**505**	**789**	**1099**	**1417**
NYT19981223.0347	T34	11KB	1KB	100	85.15	84.72	77.73	69.43	63.32	**220**	**234**	**496**	**934**	**1346**
										338	**419**	**648**	**1011**	**1625**
										18.38	20.46	25.46	31.79	40.31

d) LIN Results First Data

					Lin with different cutting point									
File Name	**Text**	**OT size**	**CT size**	**Human Ev.**	**40% cutting point**	**50% cutting point**	**60% cutting point**	**70% cutting point**	**80% cutting point**	**ERROR 40%**	**ERROR 50%**	**ERROR 60%**	**ERROR 70%**	**ERROR 40%**
APW19981019.0098	T1	8 KB	1 KB	100	68.15	55.56	40.00	27.41	20.00	**1015**	**1975**	**3600**	**5270**	**6400**
APW19981022.0269	T2	4KB	1 KB	100	78.16	65.52	56.32	43.68	32.18	**477**	**1189**	**1908**	**3172**	**4599**
APW19981030.0792	T3	5KB	1KB	100	65.65	48.09	35.88	26.72	19.08	**1180**	**2694**	**4112**	**5370**	**6547**
APW19981120.0290	T4	5KB	1KB	100	71.21	56.06	43.94	31.06	25.00	**829**	**1931**	**3143**	**4753**	**5625**
APW19981202.1274	T5	5KB	1KB	100	78.22	66.34	52.48	35.64	27.72	**474**	**1133**	**2259**	**4142**	**5224**
APW19981211.1276	T6	6KB	1KB	100	58.93	43.75	32.14	25.00	20.54	**1687**	**3164**	**4605**	**5625**	**6315**
APW19981212.0161	T7	6KB	1KB	100	61.96	46.74	30.98	22.28	14.67	**1447**	**2837**	**4764**	**6040**	**7281**
APW19981227.0870	T8	4KB	1KB	100	69.14	49.38	44.44	27.16	20.99	**953**	**2562**	**3086**	**5306**	**6243**
NYT19981001.0379	T9	20KB	1KB	100	71.79	59.19	44.08	31.49	23.68	**796**	**1665**	**3127**	**4694**	**5825**
NYT19981003.0120	T10	6KB	1KB	100	68.99	55.06	37.34	24.68	15.82	**962**	**2019**	**3926**	**5673**	**7086**
NYT19981004.0102	T11	9KB	1KB	100	66.84	56.48	44.85	31.61	23.83	**1100**	**1894**	**3042**	**4678**	**5801**
NYT19981010.0149	T12	9KB	1KB	100	70.95	63.51	54.05	34.46	22.30	**844**	**1331**	**2111**	**4296**	**6038**
NYT19981012.0334	T13	10KB	1KB	100	62.90	48.87	34.84	23.53	14.93	**1377**	**2614**	**4246**	**5848**	**7237**
NYT19981012.0359	T14	9KB	1KB	100	67.72	52.91	35.45	27.51	20.63	**1042**	**2217**	**4167**	**5254**	**6299**
NYT19981013.0354	T15	8KB	1KB	100	78.98	67.05	56.25	47.73	37.50	**442**	**1086**	**1914**	**2732**	**3906**
NYT19981013.0399	T16	8KB	1KB	100	72.67	58.39	44.72	34.16	24.22	**747**	**1732**	**3056**	**4335**	**5742**
NYT19981017.0027	T17	9KB	1KB	100	65.35	51.18	39.37	29.92	19.69	**1200**	**2383**	**3676**	**4911**	**6450**
NYT19981018.0091	T18	10KB	1KB	100	77.84	68.65	58.38	45.41	34.59	**491**	**983**	**1732**	**2981**	**4278**
NYT19981024.0050	T19	11KB	1KB	100	74.52	59.13	44.23	31.25	22.60	**649**	**1670**	**3110**	**4727**	**5991**
NYT19981104.0545	T20	11KB	1KB	100	57.55	44.34	35.38	27.83	17.45	**1802**	**3098**	**4176**	**5208**	**6814**
NYT19981105.0538	T21	8KB	1KB	100	70.62	51.03	40.21	30.93	18.04	**863**	**2398**	**3575**	**4771**	**6717**
NYT19981107.0056	T22	7Kb	1KB	100	64.12	54.12	34.12	24.71	17.65	**1288**	**2105**	**4340**	**5669**	**6782**
NYT19981107.0251	T23	5Kb	1KB	100	67.74	54.84	39.78	30.11	20.43	**1041**	**2040**	**3626**	**4885**	**6331**
NYT19981114.0079	T24	13KB	1KB	100	63.81	51.87	38.43	27.99	18.66	**1310**	**2317**	**3791**	**5186**	**6617**
NYT19981114.0129	T25	15KB	1KB	100	74.05	61.62	51.89	35.14	23.24	**673**	**1473**	**2314**	**4207**	**5892**
NYT19981121.0117	T26	7KB	1KB	100	64.90	56.29	43.71	34.44	41.72	**1232**	**1910**	**3169**	**4298**	**3396**
NYT19981122.0163	T27	9KB	1KB	100	75.26	68.56	60.31	47.94	38.66	**612**	**989**	**1575**	**2710**	**3763**
NYT19981126.0192	T28	9KB	1Kb	100	74.05	61.62	51.89	35.14	23.24	**673**	**1473**	**2314**	**4207**	**5892**
NYT19981201.0444	T29	7KB	1KB	100	56.13	39.35	26.45	19.35	9.68	**1925**	**3678**	**5409**	**6504**	**8158**
NYT19981204.0365	T30	7KB	1KB	100	66.50	58.00	42.50	28.50	20.50	**1122**	**1764**	**3306**	**5112**	**6320**
NYT19981209.0451	T31	13KB	1KB	100	75.00	63.43	52.24	33.96	26.87	**625**	**1337**	**2281**	**4362**	**5349**
NYT19981219.0117	T32	13KB	1KB	100	57.87	46.81	35.32	22.98	11.91	**1775**	**2829**	**4184**	**5932**	**7759**
NYT19981221.0377	T33	8KB	1KB	100	69.10	60.11	46.07	32.58	23.60	**955**	**1591**	**2909**	**4545**	**5838**
NYT19981223.0347	T34	11KB	1KB	100	69.87	62.01	46.29	31.00	18.34	**908**	**1443**	**2885**	**4760**	**6668**
										1015	**1986**	**3278**	**4770**	**6035**
										31.86	44.57	57.25	69.06	77.68

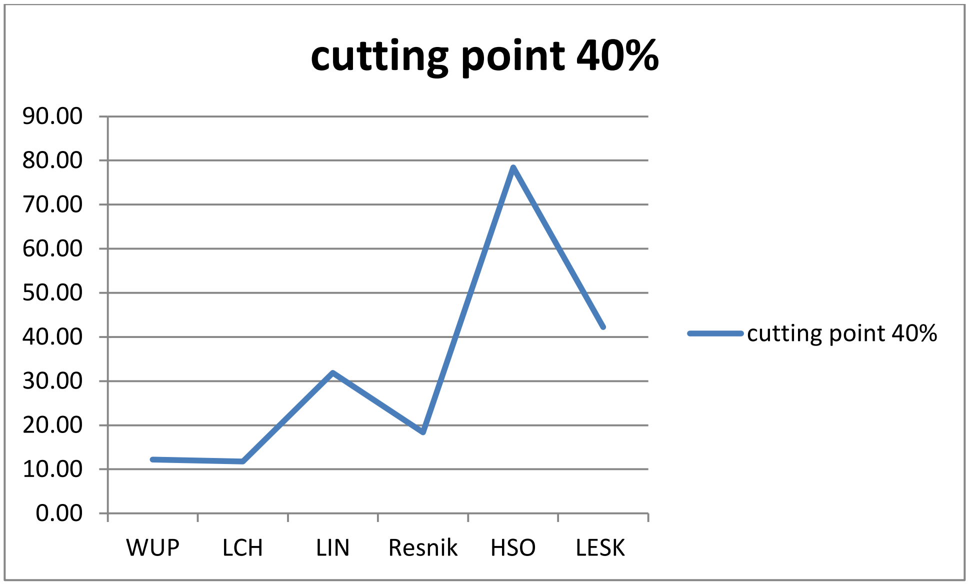
e) HSO Results First Data

File Name	Text	OT size	CT size	Human Ev.	HSO with different cutting point 40% cutting point	50% cutting point	60% cutting point	70% cutting point	80% cutting point	ERROR 40%	ERROR 50%	ERROR 60%	ERROR 70%	ERROR 80%
APW19981019.0098	T1	8 KB	1 KB	100	21.48	14.81	14.81	14.81	14.81	**6165**	**7257**	**7257**	**7257**	**7257**
APW19981022.0269	T2	4KB	1 KB	100	33.33	21.84	21.84	21.84	21.84	**4444**	**6109**	**6109**	**6109**	**6109**
APW19981030.0792	T3	5KB	1KB	100	24.43	14.50	14.50	14.50	14.50	**5711**	**7310**	**7310**	**7310**	**7310**
APW19981120.0290	T4	5KB	1KB	100	26.52	16.67	16.67	16.67	16.67	**5400**	**6944**	**6944**	**6944**	**6944**
APW19981202.1274	T5	5KB	1KB	100	30.69	9.90	9.90	9.90	9.90	**4803**	**8118**	**8118**	**8118**	**8118**
APW19981211.1276	T6	6KB	1KB	100	23.21	12.50	12.50	12.50	12.50	**5896**	**7656**	**7656**	**7656**	**7656**
APW19981212.0161	T7	6KB	1KB	100	13.04	7.07	7.07	7.07	7.07	**7561**	**8637**	**8637**	**8637**	**8637**
APW19981227.0870	T8	4KB	1KB	100	23.46	18.52	18.52	18.52	18.52	**5859**	**6639**	**6639**	**6639**	**6639**
NYT19981001.0379	T9	20KB	1KB	100	22.42	9.82	9.82	9.82	9.82	**6019**	**8132**	**8132**	**8132**	**8132**
NYT19981003.0120	T10	6KB	1KB	100	17.09	12.03	12.03	12.03	12.03	**6874**	**7740**	**7740**	**7740**	**7740**
NYT19981004.0102	T11	9KB	1KB	100	31.09	11.40	11.40	11.40	11.40	**4749**	**7850**	**7850**	**7850**	**7850**
NYT19981010.0149	T12	9KB	1KB	100	25.00	14.19	14.19	14.19	14.19	**5625**	**7363**	**7363**	**7363**	**7363**
NYT19981012.0334	T13	10KB	1KB	100	13.12	7.69	7.69	7.69	7.69	**7548**	**8521**	**8521**	**8521**	**8521**
NYT19981012.0359	T14	9KB	1KB	100	22.22	11.11	11.11	11.11	11.11	**6049**	**7901**	**7901**	**7901**	**7901**
NYT19981013.0354	T15	8KB	1KB	100	38.64	26.14	26.14	26.14	26.14	**3765**	**5456**	**5456**	**5456**	**5456**
NYT19981013.0399	T16	8KB	1KB	100	18.01	10.56	10.56	10.56	10.56	**6722**	**8000**	**8000**	**8000**	**8000**
NYT19981017.0027	T17	9KB	1KB	100	20.47	10.24	10.24	10.24	10.24	**6325**	**8058**	**8058**	**8058**	**8058**
NYT19981018.0091	T18	10KB	1KB	100	35.68	20.00	20.00	20.00	20.00	**4138**	**6400**	**6400**	**6400**	**6400**
NYT19981024.0050	T19	11KB	1KB	100	15.87	9.13	9.13	9.13	9.13	**7079**	**8257**	**8257**	**8257**	**8257**
NYT19981104.0545	T20	11KB	1KB	100	1.79	10.38	10.38	10.38	10.38	**9645**	**8032**	**8032**	**8032**	**8032**
NYT19981105.0538	T21	8KB	1KB	100	20.62	8.76	8.76	8.76	8.76	**6301**	**8324**	**8324**	**8324**	**8324**
NYT19981107.0056	T22	7Kb	1KB	100	21.76	11.18	11.18	11.18	11.18	**6121**	**7890**	**7890**	**7890**	**7890**
NYT19981107.0251	T23	5Kb	1KB	100	24.73	15.05	15.05	15.05	15.05	**5665**	**7216**	**7216**	**7216**	**7216**
NYT19981114.0079	T24	13KB	1KB	100	19.03	7.84	7.84	7.84	7.84	**6556**	**8494**	**8494**	**8494**	**8494**
NYT19981114.0129	T25	15KB	1KB	100	21.62	11.89	11.89	11.89	11.89	**6143**	**7763**	**7763**	**7763**	**7763**
NYT19981121.0117	T26	7KB	1KB	100	15.23	8.61	8.61	8.61	8.61	**7186**	**8352**	**8352**	**8352**	**8352**
NYT19981122.0163	T27	9KB	1KB	100	36.60	21.65	21.65	21.65	21.65	**4020**	**6139**	**6139**	**6139**	**6139**
NYT19981126.0192	T28	9KB	1Kb	100	21.62	11.89	11.89	11.89	11.89	**6143**	**7763**	**7763**	**7763**	**7763**
NYT19981201.0444	T29	7KB	1KB	100	9.03	3.87	3.87	3.87	3.87	**8275**	**9241**	**9241**	**9241**	**9241**
NYT19981204.0365	T30	7KB	1KB	100	19.00	11.50	11.50	11.50	11.50	**6561**	**7832**	**7832**	**7832**	**7832**
NYT19981209.0451	T31	13KB	1KB	100	29.10	15.67	15.67	15.67	15.67	**5026**	**7111**	**7111**	**7111**	**7111**
NYT19981219.0117	T32	13KB	1KB	100	14.04	3.83	3.83	3.83	3.83	**7389**	**9249**	**9249**	**9249**	**9249**
NYT19981221.0377	T33	8KB	1KB	100	20.22	9.55	9.55	9.55	9.55	**6364**	**8181**	**8181**	**8181**	**8181**
NYT19981223.0347	T34	11KB	1KB	100	15.72	3.93	3.93	3.93	3.93	**7103**	**9229**	**9229**	**9229**	**9229**
										6154	**7740**	**7740**	**7740**	**7740**
										78.45	87.98	87.98	87.98	87.98

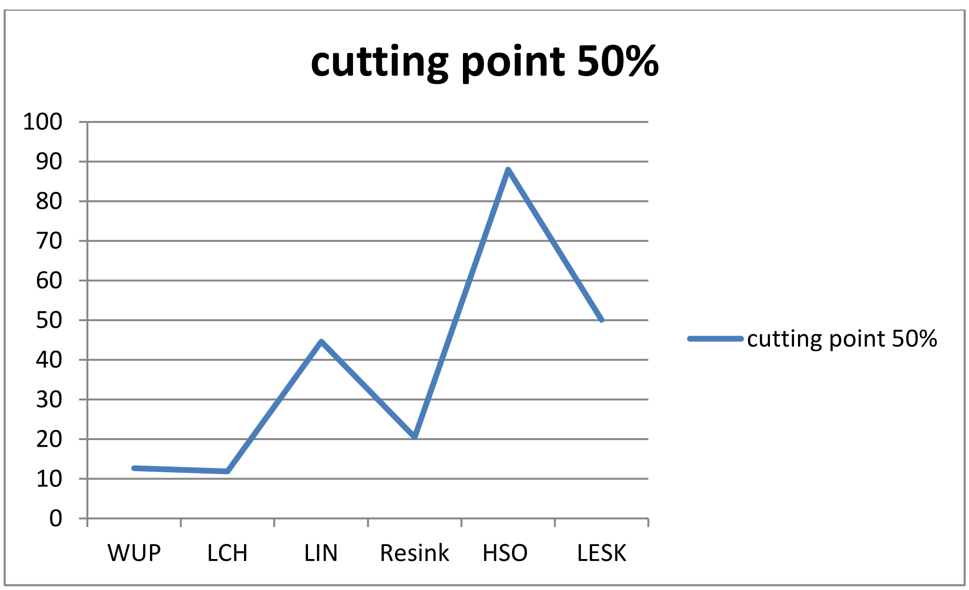
f) LESK Results First Data

					LESK with different cutting point									
File Name	Text	OT size	CT size	Human Ev.	40% cutting point	50% cutting point	60% cutting point	70% cutting point	80% cutting point	ERROR 40%	ERROR 50%	ERROR 60%	ERROR 70%	ERROR 80%
APW19981019.0098	T1	8 KB	1 KB	100	61.48	51.85	43.70	36.30	31.85	**1484**	**2318**	**3169**	**4058**	**4644**
APW19981022.0269	T2	4KB	1 KB	100	32.18	28.74	26.44	24.14	21.84	**4599**	**5079**	**5412**	**5755**	**6109**
APW19981030.0792	T3	5KB	1KB	100	43.51	37.40	32.06	29.77	22.14	**3191**	**3918**	**4616**	**4932**	**6063**
APW19981120.0290	T4	5KB	1KB	100	46.21	34.85	29.55	28.03	24.24	**2893**	**4245**	**4964**	**5180**	**5739**
APW19981202.1274	T5	5KB	1KB	100	64.36	57.43	50.50	45.54	41.58	**1270**	**1813**	**2451**	**2965**	**3412**
APW19981211.1276	T6	6KB	1KB	100	60.71	49.11	42.86	33.93	29.46	**1543**	**2590**	**3265**	**4365**	**4975**
APW19981212.0161	T7	6KB	1KB	100	71.74	61.41	53.80	46.20	40.76	**799**	**1489**	**2134**	**2895**	**3509**
APW19981227.0870	T8	4KB	1KB	100	53.09	44.44	38.27	34.57	30.86	**2201**	**3086**	**3810**	**4281**	**4780**
NYT19981001.0379	T9	20KB	1KB	100	64.48	54.41	49.62	42.07	38.29	**1261**	**2079**	**2538**	**3356**	**3808**
NYT19981003.0120	T10	6KB	1KB	100	59.49	48.73	41.14	35.44	29.75	**1641**	**2628**	**3465**	**4168**	**4936**
NYT19981004.0102	T11	9KB	1KB	100	62.69	53.37	46.39	42.49	37.82	**1392**	**2175**	**2874**	**3308**	**3866**
NYT19981010.0149	T12	9KB	1KB	100	53.38	45.95	39.86	35.81	29.73	**2174**	**2922**	**3616**	**4120**	**4938**
NYT19981012.0334	T13	10KB	1KB	100	61.99	54.75	46.15	42.53	37.56	**1445**	**2047**	**2899**	**3302**	**3899**
NYT19981012.0359	T14	9KB	1KB	100	49.21	42.86	38.62	32.28	26.98	**2580**	**3265**	**3767**	**4587**	**5331**
NYT19981013.0354	T15	8KB	1KB	100	43.75	35.80	32.39	31.25	27.27	**3164**	**4122**	**4572**	**4727**	**5289**
NYT19981013.0399	T16	8KB	1KB	100	64.60	57.14	52.17	45.34	38.51	**1253**	**1837**	**2287**	**2988**	**3781**
NYT19981017.0027	T17	9KB	1KB	100	52.76	49.61	41.73	36.22	32.28	**2232**	**2540**	**3395**	**4068**	**4586**
NYT19981018.0091	T18	10KB	1KB	100	54.59	45.41	38.38	31.35	28.11	**2062**	**2981**	**3797**	**4713**	**5168**
NYT19981024.0050	T19	11KB	1KB	100	71.15	60.10	51.44	45.19	42.79	**832**	**1592**	**2358**	**3004**	**3273**
NYT19981104.0545	T20	11KB	1KB	100	67.92	57.08	50.47	45.75	36.79	**1029**	**1843**	**2453**	**2943**	**3995**
NYT19981105.0538	T21	8KB	1KB	100	67.53	60.82	51.55	44.33	38.14	**1055**	**1535**	**2348**	**3099**	**3826**
NYT19981107.0056	T22	7Kb	1KB	100	62.35	50.59	47.65	42.94	35.88	**1417**	**2442**	**2741**	**3256**	**4111**
NYT19981107.0251	T23	5Kb	1KB	100	54.84	44.09	40.86	33.33	30.11	**2040**	**3126**	**3498**	**4444**	**4885**
NYT19981114.0079	T24	13KB	1KB	100	73.88	66.04	59.70	49.63	43.66	**682**	**1153**	**1624**	**2537**	**3175**
NYT19981114.0129	T25	15KB	1KB	100	58.38	50.27	44.86	38.92	36.22	**1732**	**2473**	**3040**	**3731**	**4068**
NYT19981121.0117	T26	7KB	1KB	100	65.56	56.29	47.68	40.40	33.77	**1186**	**1910**	**2737**	**3552**	**4386**
NYT19981122.0163	T27	9KB	1KB	100	37.63	31.96	28.87	26.80	25.77	**3890**	**4630**	**5060**	**5358**	**5510**
NYT19981126.0192	T28	9KB	1Kb	100	58.38	50.27	44.86	38.92	36.22	**1732**	**2473**	**3040**	**3731**	**4068**
NYT19981201.0444	T29	7KB	1KB	100	69.03	65.81	58.06	55.48	51.61	**959**	**1169**	**1759**	**1982**	**2341**
NYT19981204.0365	T30	7KB	1KB	100	57.00	51.50	44.50	34.50	29.50	**1849**	**2352**	**3080**	**4290**	**4970**
NYT19981209.0451	T31	13KB	1KB	100	49.63	44.40	38.43	32.84	28.73	**2537**	**3091**	**3791**	**4511**	**5079**
NYT19981219.0117	T32	13KB	1KB	100	73.19	63.40	52.77	48.09	41.70	**719**	**1339**	**2231**	**2695**	**3399**
NYT19981221.0377	T33	8KB	1KB	100	62.92	52.81	44.38	37.64	33.71	**1375**	**2227**	**3093**	**3889**	**4395**
NYT19981223.0347	T34	11KB	1KB	100	76.86	70.74	63.76	57.21	50.66	**536**	**856**	**1314**	**1831**	**2435**
										1787	**2510**	**3153**	**3783**	**4375**
										42.27	50.10	56.15	61.51	66.14

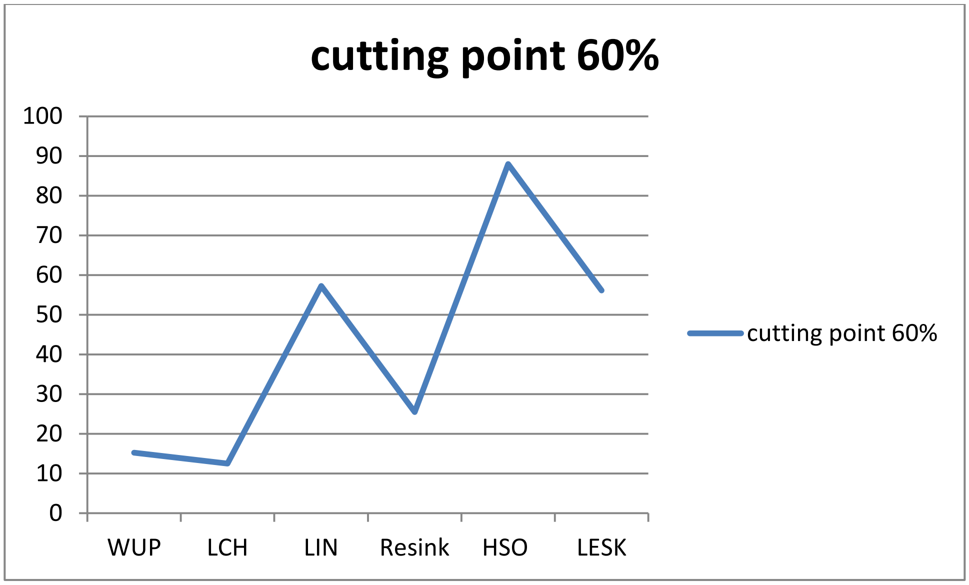
5. a) Cutting Point 40% figure



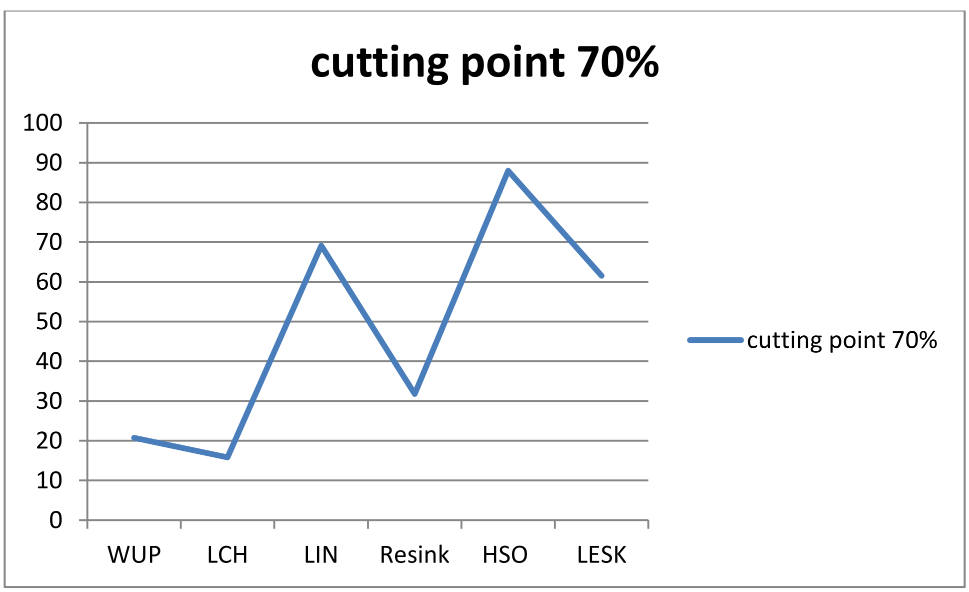
b) Cutting point 50% figure



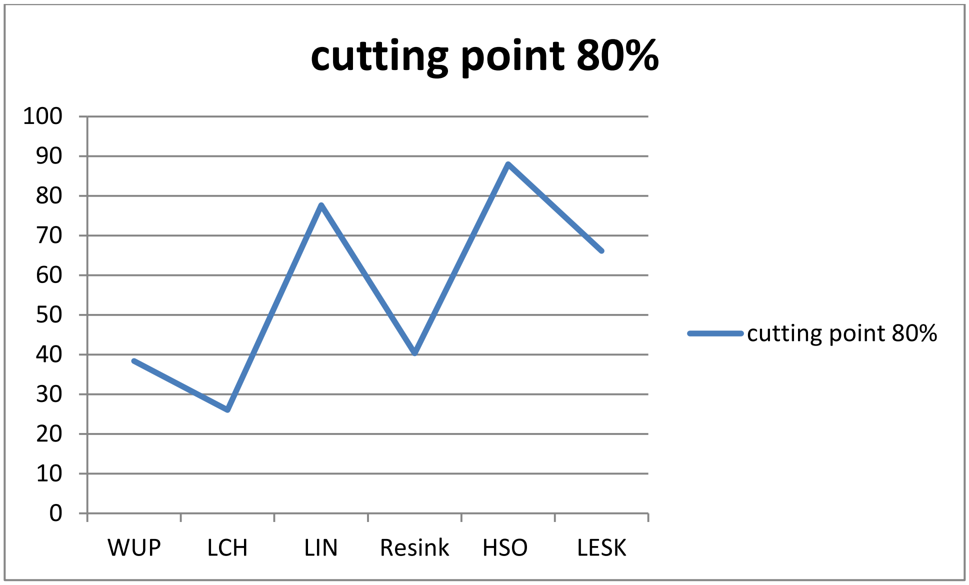
c) cutting point 60%



d) cutting point 70%



e) cutting point 80%



6. a) WuP results second data

					WuP with different cutting point									
File Name	Text	OT size	CT size	Human Ev.%	40% cutting point	50% cutting point	60% cutting point	70% cutting point	80% cutting point	ERROR 40%	ERROR 50%	ERROR 60%	ERROR 70%	ERROR 80%
Text 1	T1	5 KB	1 KB	80	84	84	82	75	63	18	18	2	21	284
Text 2	T2	11KB	1 KB	65	87	87	85	82	69	477	477	404	275	15
Text 3	T3	4KB	1KB	85	88	87	86	74	64	8	3	1	129	452
Text 4	T4	5KB	1KB	80	87	87	84	83	66	42	42	18	10	188
Text 5	T5	3KB	1KB	90	81	81	81	79	66	79	79	79	116	574
Text 6	T6	3KB	1KB	90	89	87	78	68	51	2	7	140	492	1554
Text 7	T7	6KB	1KB	80	87	87	86	78	63	55	55	34	4	289
Text 8	T8	3KB	1KB	90	78	78	76	72	62	146	146	208	321	805
Text 9	T9	3KB	1KB	90	85	85	81	77	62	27	27	81	163	783
Text 10	T10	5KB	1KB	80	91	91	90	88	68	119	119	102	58	150
Text 11	T11	6KB	1KB	75	90	88	86	81	66	218	171	129	32	73
Text 12	T12	19KB	1KB	40	90	90	89	85	74	2545	2545	2397	2005	1158
Text 13	T13	3KB	1KB	90	87	85	83	70	53	11	25	44	400	1344
Text 14	T14	5KB	1KB	75	88	83	78	76	67	156	63	12	1	63
										279	270	261	288	552
										16.70	16.43	16.15	16.96	23.50

b) LCH Results second data

LCH with different cutting point															
File Name	Text	OT size	CT size	Human Ev. %	40% cutting point	50% cutting point	60% cutting point	70% cutting point	80% cutting point	ERROR 40%	ERROR 50%	ERROR 60%	ERROR 70%	ERROR 80%
Text 1	T1	5 KB	1 KB	80	86	86	83	74	64	36	36	11	40	255
Text 2	T2	11KB	1 KB	65	86	86	84	78	68	440	440	369	160	9
Text 3	T3	4KB	1KB	85	86	86	85	81	74	1	1	0	14	120
Text 4	T4	5KB	1KB	80	87	87	84	82	75	42	42	18	4	22
Text 5	T5	3KB	1KB	90	77	77	77	74	66	160	160	160	269	574
Text 6	T6	3KB	1KB	90	87	86	85	78	55	7	14	24	140	1213
Text 7	T7	6KB	1KB	80	86	86	83	75	65	34	34	12	27	214
Text 8	T8	3KB	1KB	90	77	77	76	76	58	176	176	208	208	1015
Text 9	T9	3KB	1KB	90	84	84	77	70	59	42	42	163	415	931
Text 10	T10	5KB	1KB	80	89	89	89	86	79	86	86	86	35	2
Text 11	T11	6KB	1KB	75	89	89	89	85	69	186	186	186	105	39
Text 12	T12	19KB	1KB	40	90	90	89	86	76	2515	2515	2426	2113	1326
Text 13	T13	3KB	1KB	90	83	83	82	77	62	44	44	69	178	803
Text 14	T14	5KB	1KB	75	86	86	85	75	67	129	129	105	0	63
										278	279	274	265	471
										16.68	16.70	16.56	16.27	21.69

c) LIN Results Second Data

Lin with different cutting point														
File Name	Text	OT size	CT size	Human Ev.	40% cutting point	50% cutting point	60% cutting point	70% cutting point	80% cutting point	ERROR 40%	ERROR 50%	ERROR 60%	ERROR 70%	ERROR 40%
Text 1	T1	5 KB	1 KB	80	69	62	49	43	36	115	314	953	1370	1939
text 2	T2	11KB	1 KB	65	80	68	57	43	34	233	12	57	504	948
Text 3	T3	4KB	1KB	85	73	52	42	36	30	156	1112	1870	2375	3061
Text 4	T4	5KB	1KB	80	66	60	47	36	27	188	418	1076	1940	2813
Text 5	T5	3KB	1KB	90	64	55	42	36	25	668	1245	2351	2932	4287
Text 6	T6	3KB	1KB	90	62	54	37	28	24	780	1294	2832	3895	4338
Text 7	T7	6KB	1KB	80	77	64	54	43	28	8	263	659	1346	2750
Text 8	T8	3KB	1KB	90	57	48	42	33	20	1091	1791	2317	3300	4933
Text 9	T9	3KB	1KB	90	68	59	49	42	29	469	931	1651	2326	3707
Text 10	T10	5KB	1KB	80	72	58	38	26	15	66	491	1763	2957	4241
Text 11	T11	6KB	1KB	75	68	58	49	39	31	55	291	683	1281	1914
Text 12	T12	19KB	1KB	40	81	73	58	44	33	1697	1118	321	13	56
Text 13	T13	3KB	1KB	90	62	53	32	23	18	803	1344	3403	4444	5136
Text 14	T14	5KB	1KB	75	68	65	59	49	42	46	105	253	683	1086
										455	766	1442	2098	2943
										21.34	27.68	37.98	45.80	54.25

d) Resnik Results Second Data

				Resnik with different cutting point										
File Name	Text	OT size	CT size	Human Ev.	40% cutting point	50% cutting point	60% cutting point	70% cutting point	80% cutting point	ERROR 40%	ERROR 50%	ERROR 60%	ERROR 70%	ERROR 80
Text 1	T1	5 KB	1 KB	80	80	75	68	64	57	0	30	134	255	528
Text 2	T2	11KB	1 KB	65	80	78	75	66	60	233	171	100	2	29
Text 3	T3	4KB	1KB	85	85	82	70	57	53	0	7	215	776	1040
Text 4	T4	5KB	1KB	80	75	75	74	70	63	22	22	34	107	292
Text 5	T5	3KB	1KB	90	70	70	64	57	47	408	408	668	1115	1834
Text 6	T6	6KB	1KB	90	86	80	70	66	52	14	91	395	599	1465
Text 7	T7	6KB	1KB	80	81	81	76	72	61	1	1	19	70	345
Text 8	T8	3KB	1KB	90	74	71	69	57	45	243	364	458	1091	1994
Text 9	T9	3KB	1KB	90	84	78	75	71	53	42	133	235	365	1357
Text 10	T10	5KB	1KB	80	85	85	82	66	54	26	26	3	193	691
Text 11	T11	6KB	1KB	75	81	80	76	66	57	39	26	0	83	310
Text 12	T12	19KB	1KB	40	89	86	79	74	68	2367	2141	1506	1158	787
Text 13	T13	3KB	1KB	90	83	83	72	70	62	44	44	336	400	803
Text 14	T14	5KB	1KB	75	74	70	68	61	52	1	21	46	186	517
										246	249	296	457	857
										15.68	15.77	17.22	21.38	29.27

e) LESK Results Second Data

				LESK with different cutting point										
File Name	Text	OT size	CT size	Human Ev.	40% cutting point	50% cutting point	60% cutting point	70% cutting point	80% cutting point	ERROR 40%	ERROR 50%	ERROR 60%	ERROR 70%	ERROR 80%
Text 1	T1	5 KB	1 KB	80	36	29	27	26	22	1939	2606	2789	2882	3372
Text 2	T2	11KB	1 KB	65	50	44	39	31	28	225	447	674	1146	1396
Text 3	T3	4KB	1KB	85	37	30	23	22	22	2269	3061	3834	3972	3972
Text 4	T4	5KB	1KB	80	65	51	39	31	29	220	867	1654	2356	2579
Text 5	T5	3KB	1KB	90	38	32	26	25	25	2732	3355	4043	4287	4287
Text 6	T6	3KB	1KB	90	33	25	23	22	22	3211	4188	4491	4646	4646
Text 7	T7	6KB	1KB	80	39	35	32	29	28	1651	2057	2277	2587	2668
Text 8	T8	3KB	1KB	90	27	26	23	23	23	4001	4150	4455	4455	4455
Text 9	T9	3KB	1KB	90	38	33	32	29	28	2707	3259	3405	3707	3863
Text 10	T10	5KB	1KB	80	62	53	40	32	28	325	735	1627	2282	2694
Text 11	T11	6KB	1KB	75	54	47	40	35	30	442	807	1201	1627	2066
Text 12	T12	19KB	1KB	40	54	48	41	36	32	205	65	1	17	70
Text 13	T13	3KB	1KB	90	67	60	45	38	33	544	900	2025	2669	3211
Text 14	T14	5KB	1KB	75	38	34	33	33	33	1406	1674	1768	1768	1768
										1563	2012	2446	2743	2932
										39.53	44.86	49.46	52.37	54.15

f) HSO Results Second Data

				HSO with different cutting point										
File Name	Text	OT size	CT size	Human Ev.	40% cutting point	50% cutting point	60% cutting point	70% cutting point	80% cutting point	ERROR 40%	ERROR 50%	ERROR 60%	ERROR 70%	ERROR 80%
Text 1	T1	5 KB	1 KB	80	36	25	25	25	25	1939	2977	2977	2977	2977
Text 2	T2	11KB	1 KB	65	39	19	19	19	19	674	2129	2129	2129	2129
Text 3	T3	4KB	1KB	85	29	20	20	20	20	3184	4254	4254	4254	4254
Text 4	T4	5KB	1KB	80	29	19	19	19	19	2579	3709	3709	3709	3709
Text 5	T5	3KB	1KB	90	38	25	25	25	25	2732	4287	4287	4287	4287
Text 6	T6	3KB	1KB	90	18	17	17	17	17	5128	5294	5294	5294	5294
Text 7	T7	6KB	1KB	80	43	24	24	24	24	1346	3178	3178	3178	3178
Text 8	T8	3KB	1KB	90	31	17	17	17	17	3435	5265	5265	5265	5265
Text 9	T9	3KB	1KB	90	39	27	27	27	27	2577	4022	4022	4022	4022
Text 10	T10	5KB	1KB	80	21	15	15	15	15	3424	4241	4241	4241	4241
Text 11	T11	6KB	1KB	75	34	21	21	21	21	1720	2914	2914	2914	2914
Text 12	T12	19KB	1KB	40	36	18	18	18	18	17	475	475	475	475
Text 13	T13	3KB	1KB	90	22	15	15	15	15	4669	5625	5625	5625	5625
Text 14	T14	5KB	1KB	75	45	34	34	34	34	873	1674	1674	1674	1674
										2450	3574	3574	3574	3574
										49.50	59.79	59.79	59.79	59.79