Extracting instances to improve semantic matching in the domain of Renewable Energy

(استخراج الحالات لتحسين المطابقة الدلالي في مجال الطاقة المتجددة)

By

Sadam Hamdan Ahmed

THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR MASTER DEGREE

IN

COMPUTER INFORMATION SYSTEM

SUPERVISOR

Dr. Ahmad K. A. Kayed

DEPARTMENT OF COMPUTER INFORMATION SYSYTEM

FACULITY OF INFORMATION TECHNOLOGY

MIDDLE EAST UNIVERSITY

Amman, Jordan

2014

Examination Committee Decision

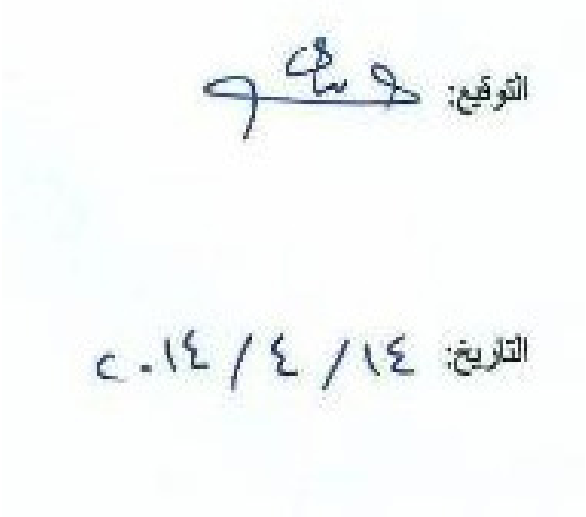
This is to certify that the thesis en titled "Extracting Instances to Improve Semantic Matching in the Domain of Renewable Energy" was successfully defended and approved on 14/4/2014.

Examination Committee Member	Signature
1- Dr. Ahmad Kayed	
2- Dr Rdwan Abd Jassar	
3- Dr. Ali Assad Ahmed	

أ.د. علي أسعد الداود
Prof. Ali As'ad Al-Daoud

اقرار التفويض

أنا صدام حمدان أحمد، أفوض جامعة الشرق الأوسط بتزويد نسخ من رسالتي للمكتبات أو المؤسسات أو الهيئات أو الأفراد عند طلبها.



Authorization Statement

I'm Sadam Hamdan Ahmed, authorize the Middle East University to supply a copy of my thesis to libraries, establishments or individuals upon their request.

Signature: 

Date: 14/4/2014

DEDICATION

To the light of my life …

My Mother,

My Father,

And My Sisters.

ACKNOWLEDGMENTS

I would like to express my sincere appreciation to Dr. Ahmad Kayed for his guidance, helping and encouraging my efforts during this research and support and motivation throughout Master's Thesis.

I owe more than thanks to my family members which includes my parents and my sisters for their financial support and encouragement, I'm very grateful to them.

I would like to thank my family. Her support, encouragement, quiet patience and unwavering love and were undeniably. They were always standing with me through the good times and bad.

Finally, for their constant love, support, and remembering to feed me when I couldn't remember to feed myself.

Thank You All..

ABSTRACT

Renewable energy (RE) has many resources, and the increased usage of these resources led to the spread of renewable energy applications. These applications and services are provided by different providers. Those providers use different architectures for their RE systems which are considered not compatible. On the other hand, the providers have many ways to describe their services which considered as a big challenge that may face the customers. This research explores the possibility to automate an environment whereby the customers can find the appropriate services provider that meets their requirements. Moreover, this research presents ontology in the domain of RE. This research explores the possibility to measure the distance between the RE providers and customer requirements. We developed a model to improve semantic matching results by using instances in the domain of RE.

Our model consists three stages as follow: (1) Extracting instances of RE providers (2) Extracting instances of customer requirements (3) Matching process among both extracted instances. The matching process has done by using semantic similarity measure. This model suggests an appropriate RE providers that meets customer requirements. Also during our study, we collected many documents and reports that discussed RE providers and consumer quires to collection their data related to each of them. Then convert these data into text file in order to extract their instances.

The proposed model is evaluated by comparing the results of our approach with the results of traditional approaches. The experiments have conducted in order to check the efficiency of the proposed model. However, the results demonstrated that using instances reduced the error to (10%).

Keywords: Renewable Energy, Ontology, Ontology Concept, Extracting Instances, Semantic Matching.

الخلاصة

لدى الطاقة المتجددة العديد من الموارد أضافةً الى زيادة استخدام هذه الطاقة السبب الذي أدى إلى انتشار تطبيقات الطاقة المتجددة. وتقدم هذه التطبيقات والخدمات من قبل مختلف مزودي الخدمات. يستخدم هؤلاء المزودين بناءات مختلفة من انظمة الطاقة المتجددة التي تعتبر غير متكاملة. على صعيد اخر, لدى مزودي الطاقة المتجددة طرق مختلفة لوصف خدماتهم التي تعتبر تحدي كبير يواجه الزبائن. يقوم هذا البحث باستكشاف امكانية اتمتة البيئة اينما استطاع الزبائن الوصول الى مزود الخدمات المناسب الذي يلبي متطلباتهم واحتياجاتهم.

إضافة لذلك، قدم هذا البحث استخدام الأنتولوجيا في مجال الطاقة المتجددة. ان في أطروحتنا هذه نستعرض أمكانية قياس التشابه بين مزودي الطاقة المتجددة و متطلبات المستخدمين. فقد طورنا نموذج لتحسين نتائج التشابه الدلالي بأستخدام الحالات في مجال الطاقة المتجددة.

هذا النموذج يتكون من ثلاثة أقسام: (1) استخراج الحالات من مزودي الطاقة المتجددة (2) استخراج الحالات من متطلبات المستخدمين (3) عملية التطابق بين الحالات المستخرجة. استخدمنا قياس التشابه الدلالي في عملية التطابق. يقوم هذا النموذج لاقتراح افضل مزود للطاقة المتجددة الذي يلبي متطلبات العملاء. وخلال بحثنا هذا جمعنا ودرسنا العديد من المستندات والتقارير التي ناقشت مزودي الطاقة المتجددة واستعلامات المستخدمين من خلال جمع البيانات المتعلقة بكل واحد منهم ثم تحوليها الى ملف نصي لأستخراج الحالات.

تم تقييم اداء النموذج بطريقة مقارنة نتائج الطريقة المقترحة مع نتائج الطرق التقليدية. وقد تم اجراء التجارب من اجل فحص كفاءة النظام المقترح واشارت نتائج الدراسة من خلال استخدام الحالات الى تخفيض نسبة الخطأ الى (10%).

كلمات البحث: الطاقة المتجددة, علم الوجود, مفاهيم علم الوجود, استخراج الحالات, التشابه الدلالي.

TABLE OF CONTENTS

Examination committee decision……..……..II

اقرار التفويض……..……..……III

Authorization Statement...……..……IV

Dedication……....V

Acknowledgment……..……..…VI

Abstract……..……..…...VII

Abstract in Arabic…….……..…..VIII

Table of Contents……..…….IX

List of Figures……..……..XII

List of Tables..……..……XIII

Abbreviations…….……..……..XV

CHAPTER 1 (Introduction)……...……1

1.1 Overview……..……..……1

1.2 Renewable Energy……..……..……1

1.3 Renewable Energy Ontology……2

1.4 Semantic Similarity……...……...…..3

1.5 Extracting Instances……..……...……..…….3

1.6 The Problem Definition……..……...….4

1.7 Motivation……...……..…….5

1.8 Objective of the Thesis……...……....5

1.9 Methodology……...……....6

1.10 Organization of the Thesis……...……..……..8

CHAPTER 2 (Literature Review & Related Works)…….....9

2.1 Introduction……..9

2.2. Literature Review……..……..…....9

2.2.1 Building Ontology…….……..…….....9

2.2.2 Extracting Instances…….10

2.2.3 Semantic Matching……..……13

CHAPTER 3 (Renewable energy Ontology and Similarity Measures) ……..... 16

3.1 Introduction…….……16

3.2 Building Ontology……..…….16

3.3 Extracting Renewable Energy Ontology Concepts……...…..17

3.4 Semantic Matching Measures…….22

3.4.1 Our Experiment…….23

CHAPTER 4 (Methodology and Proposed Model) 26

4.1 Overview…….…….……...…….26

4.2 The Original Instances-based Data Extraction…….…….26

4.2.1 Modify IDE algorithm for Extracting Instances.…….……27

4.2.2 Applying the Modified Algorithm……...…28

4.3 Tested the Proposed Work....……...……..…..30

4.3.1 Extracting Instances from Providers and Queries……..….31

4.4 Decision Making.…….……...……37

CHAPTER 5 (Experimental Results)…..……..……....….42

5.1 Evaluation Methods…….……..….42

5.2 Semantic Matching Methods……..….42

5.2.1 Semantic Matching with KAON…….……..44

5.2.2 Semantic Matching with (WUP >= 0.5)……...44

5.2.3 Semantic Matching with (WUP >= 0.6)……45

5.2.4 Semantic Matching with (WUP >= 0.7)……..…….....46

5.2.5 Semantic Matching with (WUP >= 0.8)……....46

5.2.6 Semantic Matching with (WUP >= 0.9)……..……..47

5.3 Semantic Matching with Frequency Normalization……...……..……..48

5.3.1 Semantic Matching with KAON…….50

5.3.2 Semantic Matching with (WUP >= 0.5)……51

5.3.3 Semantic Matching with (WUP >= 0.6)……51

5.3.4 Semantic Matching with (WUP >= 0.7)……52

5.3.5 Semantic Matching with (WUP >= 0.8)……53

5.3.6 Semantic Matching with (WUP >= 0.9)……53

5.4 Analysis of the Results……54

5.5 Comparison with Related Work……56

5.6 Main Contribution……57

CHAPTER 6 (Conclusion and Future Work)……57

6.1 Overview……57

6.2 Conclusion……57

6.3 Summary……58

6.4 Future Work……58

References……60

Websites References……69

Appendix……70

LIST OF TABLES

Table 3-1 : Results of extracting ontology concepts for RE……20

Table 3-2 : Results of extracting ontology concepts for RE……....20

Table 3-3 : Standard concepts similarity measures……..…24

Table 4-1: Sample of extracted instances manually...……..29

Table 4-2: Sample of extracting instances……...30

Table 4-3: Sample of semantic matching……...…….34

Table 4-4: Sample of expert human results…….……..……..35

Table 5-1: Average error for semantic matching…….……....…43

Table 5-2: Presents sample of the results of semantic matching with KAON…….44

Table 5-3: Presents sample of the results of semantic matching with (WUP >= 0.5)…….45

Table 5-4: Presents sample of the results of semantic matching with (WUP >= 0.6)…….....45

Table 5-5: Presents sample of the results of semantic matching with (WUP >= 0.7)…….....46

Table 5.6: Presents sample of the results of semantic matching (WUP >= 0.8) ……..…46

Table 5-7: Presents sample of the results of semantic matching (WUP >= 0.9)……..47

Table 5-8: presents the overall error for all experiments…….…….48

Table 5-9 presents sample of results for the frequency per average……..……...49

Table 5-10: Presents sample of the results of semantic matching with KAON……50

Table 5-11: Presents sample of the results of semantic matching with (WUP >= 0.5)…….51

Table 5-11: Presents sample of the results of semantic matching with (WUP >= 0.6)…….51

Table 5-12: Presents sample of the results of semantic matching with (WUP >= 0.7)…….52

Table 5-13: Presents sample of the results of semantic matching with (WUP >= 0.8)…….53

Table 5-14: Presents sample of the results of semantic matching with (WUP >= 0.9)……..…53

LIST OF FIGURES

Figure 3-1 : The front-end of the KAON TextToOnto Tool……....18

Figure 3-2 : Create new corpus function using KAON TextToOnto Tools……..…19

Figure 3-3 : New Term Extraction function using KAON TextToOnto Tool…..……...19

Figure 3-4: Concepts extraction for multi frequency…….……..21

Figure 4-1: User Interface for proposed application.....28

Figure 4-2: Flow chart of match percentage calculation and decision making……....39

Figure 4-3: Main block diagram of the presented thesis ……..…41

LIST OF ABBRIVATIONS

RE	:	Renewable Energy.
IT	:	Information Technology.
RDF	:	Resource Description Framework.
KAON	:	KArlsuhe ONtology.
ASIA	:	Automatic Set Instances Acquirer.
WS4J	:	WordNet Similarity for Java.
WuP	:	Wu-Palmer.
UMLS	:	Unified Medical Language System.
LCS	:	Least Common Sub-summer.
IDE	:	Instances-based Data Extraction.
API	:	Application Programming Interface.

CHAPTER 1

Introduction

1.1Overview

This chapter provides a background on the four main concepts covered by this research, namely: renewable energy, renewable energy ontology, semantic similarity and instances. This chapter indicates the reason of selecting RE domain. Moreover, this chapter investigates the research problem as well as the problem solution. Also, this chapter shows the motivation of the research which gives attention to the research problem and how it has been addressed. The objectives of the research that must be achieved are mentioned in this chapter. We illustrate the research significant and give an idea about our own contributions.

1.2 Renewable Energy

Jacobson and Delucchi illustrates that the increased usage of renewable energy resources led to the spread of renewable energy (RE) applications. These applications need information technology and computing support (Jacobson, M. and Delucchi, M. 2011). Also Al-Bahadili stated that applications in the RE area may include modeling and simulations, diagnosis, control, and monitoring of renewable energy process and systems (Al-Bahadili H. et al. 2013). The author stated that the IT can play a main role in developing efficient and reduce cost of RE. Computing powers will solve many issues in renewable energy domain such as imprecision, uncertainty, low cost, and robust (Al-Bahadili H. et al. 2013).

1.3 Renewable Energy Ontology

There are many definitions for ontology such as Gruber defined ontology is a formal, explicit specification of a shared conceptualization (Gruber T. 1993). Zhang et al. stated that ontology provides set of concepts and their interrelationships in a specific domain to facilitate understanding and automatic processing of text (Zhang J. et al. 2012). Also, Jiang defined ontology as an abstract description system for knowledge composition in a certain domain, organizing concepts (terms) in a domain in a hierarchical way and describing relationships between terms using a small number of relational descriptors, an ontology supplies a standardized vocabulary for representing entities in the domain (Jiang R. 2013).

In order to automate an environment whereby the consumers can reach to the appropriate RE service providers, an ontology based semantic matching are needed for precise discovering and provisioning computing resources across various RE providers.

In ontology building process, they are many ways to identify the main concepts in a domain. To build a new ontology, it is important to 1) identify the domain of ontology, why the ontology is being built, what its intended uses. 2) Knowing sources: documents, experts and existing ontologies. and 3) Building ontology using a suitable ontology building tool (KAON, Protégé, etc.) (Fernández-López M. 1999) (Bermejo J. 2007). Kayed et al. illustrated some ontology's building tools (such as KAON, protégé, Neon, etc.). They assumed that extracting ontology concepts that consisted of concepts that are not only the most frequent terms but also those having high ontological relevance keywords (Kayed A. et al. 2010).

1.4 Semantic Similarity

Nagwani explained that the semantic similarity is a concept whereby a set of documents or terms within term lists assign a metric based on the likeness of their meaning / semantic content. Various semantic similarity techniques are available which can be used for measuring the semantic similarity between text documents (Nagwani N. 2011). This thesis explores the possibility to improve matching based semantic similarity measure between the RE providers and customer requirements.

1.5 Extracting Instances

The main aim of this research is to enhance the matching between RE providers and consumers' requirements. Enhance the matching process may include enhancing the keywords matching by using semantic matching. Also Al-Bahadili et al. illustrated that the semantic matching may be improved by retrieving information that are more precise and by satisfies consumers' intentions (Al-Bahadili H. et al. 2013). This research will use instances that will help in enhancing the semantic matching. It is difficult to extract instances from various data sources manually, so we need tools that can extract instances automatically.

There are several ways to extract instances from texts such as: the method which is proposed by Anantharangachar, et al. 2013. In their approach, they are populating an existing ontology with instance information present in the natural language text provided as input. They demonstrate heuristics to extract information from the unstructured text and for adding it as structured information to the selected ontology. This identification of the relevant ontology is critical, as it is used in identifying relevant information in the text. They extract information in the form of semantic triples from the text, guided by the concepts in the ontology. Then convert the extracted information

about the semantic class instances into Resource Description Framework (RDF) and append it to the existing domain ontology.

Also De Boer et al. 2006 presented a generic approach for achieve extract information from documents which are available on the World Wide Web (WWW). The previous method exploits redundancy of information to compensate for loss of precision caused by the use of domain independent extraction methods, and presented an outline for this approach in the form of a framework that is applicable in various domains and described the prerequisites of this approach. This research will extract instances in the RE domain which will help to enhance and improve the semantic matching in this domain.

1.6 The Problem Definition

The main problem of this research is semantic matching and how we can improve it to solve some semantic issues such as too many matching or too specific ones, thus; this research faced many problems which can be summarized in the following points:

1- How can we build ontology in the domain of RE?
2- How can this ontology serve the customer to match their requirements by using semantic matching techniques?
3- How we can extract instances?
4- How the instance can improve the semantic matching?

1.7 Motivation

Al- Bahadili et al. stated that the renewable energy expanding day by day due to its advantages to small and medium businesses organizations by assisting them to focus on their core business competencies. Renewable energy provide effective and affordable (Al-Bahadili H. et al. 2013).

IT tools helping organizations to investee a new infrastructure, licensing new software without spending high cost on in-house resources and technical equipment. It is essential to have an intelligent mechanism, which can understand and analysis consumers' query semantically. There are several search engines to get information which are adopting keywords matching mechanism. Keywords' matching is insufficient due to the retrieval of a large amount of irrelevant information, which has shortness in understanding consumers' query intentions (Haboush A. 2014). Dai et al. stated that the ontology allows semantic analysis of consumers' queries and a heuristic search, the expected information can be retrieved more precisely and completely that's satisfying consumers' intentions (Dai W. et al. 2011). This motivates this research to find some issues for RE, we are looking to:

1- Identify the ability of extracting instances in RE domain.
2- Find new tools to enhance the result of semantic matching.
3- Using instances that contributed in ontological techniques to find the way that meets customer requirements in this domain.

1.8 Objective Of The Thesis

The main objective of this work is to improve matching to be more accurate and relevant to customers' requirements. This matching is semantic, not keywords matching. Semantic matching must be improved by finding a way to help improve the matching results and to access to relevant matching.

This research identified the role of instances in improving the semantic matching in the domain of RE and using it to solve some semantic issues such as too many matching or too specific which is too small matching. This research explained a modification of an existing algorithm named Instances-based Data Extraction (IDE) that help to improve and enhance the semantic matching**.** This will contribute in create ontology in the domain of RE. It also serve and enhance the process of choosing RE and how we can generalize it to any other domain. The research developed and deployed ontology framework for the renewable energy. Additionally, number of issues have been addressed in this research such as:

1- The extracting of instances in the domain of RE.
2- The implementation and evaluation of the proposed ontology.
3- How can the ontology helps customer to decide which renewable energy provider is meets their requirements using the extracted instances.

1.9 Methodology

The main aim of this methodology is to used semantic analysis to meet the providers services' and customers requirements' in a way that facilitate the process of finding the best Renewable Energy. Additionally, that has been used to developed our model contains the following phases:

- Study and Analysis Phase.

- Design and Implementation Phase

- Evaluation Phase.

Study and Analysis Phase

The first step in the studying and analysis has been the collection of data, acquiring information and knowledge for the following:

1- Collected data related to RE domain.
2- Checking if there is any RE ontology to enhance it, if no ontology exists we will construct new one using KAON ontology's building tool.
3- Extracting instances to help to achieving matching and get best and relevant results.

We will explain the full details for this phase in the chapter three**.**

Design and Implementation Phase

In this phase we designed a model and implement it with the necessary collection data and information in order to extract instances to improve semantic matching in the domain of renewable energy. We discussed this phase in details in the chapter four.

Evaluation Phase

This phase consists testing our model to measure how instances improving the semantic matching. For more details see chapter five.

1.10 Organization of the Thesis

Chapter 1: Presents a theoretical background about the RE, its ontology, semantic similarity matching, how can we improve semantic matching, and how we can extract instances.

Chapter 2: Illustrates the process of applying the effectiveness measure extract instances in improving semantic matching.

Chapter 3: Explain the process of extracting ontology concepts in details and how to choose the semantic similarity measures.

Chapter 4: Display in detail the component of our model and matching process in details.

Chapter 5: Shows the experimental results for our proposed model.

Chapter 6: States conclusion and future works.

CHAPTER 2

Literature Review & Related Works

2.1 Introduction

This chapter gives a brief idea about the most relevant work in the literature that relates to our study. We provide a background and literature review of the three main concepts covered by this research, namely: building ontology, extract instances and semantic matching.

2.2 Literature Review

This part investigation of existing study and research which is relevant to our theme and present some background reading required to give context to our research.

2.2.1 Building Ontology

Corcho reviewed and compared the main characteristics and guidelines that help selecting the most appropriate methodologies, tools and languages which can help for building ontology in a specific domain, as well as the main relationships among them (Corcho O. et al. 2003).

Kayed demonstrated several experiments to extract concepts to build ontologies that improve the description process for software components embedded in a web document. The auther built ontology (mainly concepts) for some software components then used them to solve some semantic problems. Then, collected many documents that describe components in Net and Java from several and different resources. Concepts were extracted and used to decide which domain of any given description (semantic) is close or belong to (Kayed A. 2011).

Kayed et al. built ontologies in the domain of software components (.Net and Java.). Used KAON1 (KArlsruhe ONtology) to extract concepts to build ontologies (Kayed A. et al. 2011).

Noy and McGuiness presented a guide that helps the developer to build ontology. First, the authors defined the concept of ontology. They presented a guide that has instructions about the ways to build ontology. And about the reason of building ontology, the authors claimed that ontology represents a common vocabulary about the searches of the users. Also, developers build the ontology to share common understanding of the structure of information among people or software agents. Besides that, authors suggest that the other reason of building ontology is to enable users and customer to reuse the knowledge and analyze domain knowledge (Noy and McGuinness 2009).

Rajan developed an ontology-based search engine that makes use of semantic matching methodology and performs accurate retrieval of Web services in Vaccination of Healthcare domain (Rajan J. and Lakshmi M. 2012).

Swartout described how they have used a large-scale ontology develop to a specialized, domain-specific ontology semi automatically and discuss the relation between ontologies and the process of developing a system (Swartout B et al. 1996).

2.2.2 Extracting Instances

Bhavana et al. described an open-domain information extraction method for extracting concept-instance pairs from an HTML corpus. Their method relies on a novel approach for clustering terms found in HTML tables, and then assigning concept names to these clusters. The method can be efficiently applied to a large corpus (Bhavana D. et al. 2011).

Bo et al. proposed a method for extracting information of crop diseases on Chinese web pages. Employ an ontology-based way to implement information extraction from the content blocks. A top-down method is adopted to construct the ontology of crop diseases. In the extraction process, the concepts, relations and instances of ontology is used to extract the entities. The event is extracted by an optimal classification of paragraph groups in a content block (Bo J. et al. 2013).

Myo-Myo et al. investigated the problem of extracting link information of relationship instances from a web site. They define the notion of link chain and formulate the link chain extraction problem. An extraction method based on sequential covering has been proposed to solve the problem (Myo-Myo N. et al. 2003).

Richard and William presented a system named ASIA (Automatic Set Instance Acquirer), which takes in the name of a semantic class as input (e.g., “car makers”) and automatically outputs its instances (e.g., “ford”, “nissan”, “toyota”). ASIA is based on recent advances in web based set expansion the problem of finding all instances of a set given a small number of “seed” instances. This approach effectively exploits web resources and can be easily adapted to different languages. In brief using language dependent hyponym patterns to find a noisy set of initial seeds, and then use a state-of-the-art language-independent set expansion system to expand these seeds. The proposed approach matches or outperforms prior systems on several English language benchmarks (Richard C. and William W. 2008).

Raghu et al. described an approach to populate an existing ontology with instance information present in the natural language text provided as input. An ontology is defined as an explicit conceptualization of a shared domain. This approach started with a list of relevant domain ontologies created by human experts, and techniques for identifying the most appropriate ontology to be extended with information from a given text. Then demonstrate heuristics to extract information from the unstructured text and for adding it as structured information to the selected ontology. This

identification of the relevant ontology is critical, as it is used in identifying relevant information in the text. They extracted information in the form of semantic triples from the text, guided by the concepts in the ontology. then convert the extracted information about the semantic class instances into Resource Description Framework (RDF) and append it to the existing domain ontology (Raghu A. et al. 2013).

Sui et al. put forward a weakly-supervised method which can synchronously extract instances and attributes for a concept based on web information. Firstly, they automatically generated and evaluated the contextual patterns in which instances and attributes co-occur on the Web. Secondly, they extracted candidate instances and attributes using the patterns extracted above, and evaluate them with two methods. (1) based on the associations between instances and attributes using determinate instances (i.e. the seed instances) to evaluate the accuracy of the candidate attributes; and used the determinate attributes (i.e. the seed attributes) to evaluate the accuracy of the candidate instances. (2) they used the contextual distribution similarity to evaluate the accuracy of instance extraction and attribute extracting (Sui Z. et al. 2012).

Vadrevu et al. presented automated techniques for extracting metadata instance information by organizing and mining a set of news Web sites. They developed algorithms that detect and utilize HTML regularities in the Web documents to turn them into hierarchical semantic structures encoded as XML. They presented tree-mining algorithms that identify key domain concepts and their taxonomical relationships, also extracted semi-structured concept instances annotated with their labels whenever they are available (Vadrevu S. et al. 2004).

De bore et al. described to a specific subtask of ontology population, the extraction of instances of relations. Presented a generic approach with which is able to extract information from documents on the Web. The method exploits redundancy of information to compensate for loss of precision caused by the use of domain independent extraction methods. They presented the general approach and

describe implementation for a specific relation instance extraction task in the art domain (De Boer V. et al. 2006).

2.2.3 Semantic Matching

Budanitsky proposed to evaluate similarity measurements based on WordNet. However, the authors had evaluated five measurements lexical semantics distance. The authors mentioned that most of their work was limited to the narrower notion of similarity measures. These relationships include not just hyponymy and the nonhyponymy relationships in WordNet such as meronymy but also associative and ad hoc relationships. As the authors mentioned, these can include just about any kind of functional relation or frequent association in the world (Budanitsky 2006).

Fazzinga et al. expected that adding semantics to web data will make huge change in web information technology which is the next step in development of information technology. However, they discussed building new approach based on novel methods in order to semantic web search. Semantic web search allows for semantic processing that applied on semantic queries and evaluating complex web structure. Authors have added ontological structure and semantics to web pages that allows for attaching a meaning to web search queries and web pages. Also, semantic search allows for formulating and processing ontology based on complex search queries. They claimed that, they have obtained a general or vertical semantic of web search interface. They have addressed the web into an ontological knowledge base. Furthermore, they have built a formal model behind their approach and developed an implementation for this model on desktop search. To approve their solution model they have made some of tests based on semantic web search for movie database. After making split operation on the text each word is stored in specific location in special array, and each element in array will be processed individually. System takes each element in the array and compares it with system lexicon by comparing words with dictionaries according to the type of word: verb, adverb, noun,…etc (Fazzinga et al. 2011).

Li et al. highlighted that the similarity between words and concepts had become difficult problem that face many applications and artificial intelligence. They tried to predict the determination of semantic similarity by a number of information resources that contain semantic information from lexical taxonomy. They also indicated how information sources could be used effectively by using variety of strategies for using various possible information resources. However, authors argued that all first-hand information sources need to be processed in similarity measure. Besides that, authors claimed that humans can compare word similarity with a finite interval between similar and non-similar (Li et al. 2003).

Lukasiewicz et al. presented a novel approach to Semantic Web search, which allows for a semantic processing of Web search queries relative to an underlying ontology, and for evaluating ontology-based complex Web search queries that involve reasoning over the Web (Lukasiewicz T. et al. 2012).

Michelizzi described the different types of the semantic similarity: 1) Path Length Similarity Measures: a- Path b- Wu & Palmer c- Leacock & Chodorow 2) Leacock & Chodorow 2) Information Content Similarity Measures: a- Resnik b- Lin c- Jiang & Conrath 3) Semantic Relatedness Measures a- Extended Gloss Overlaps (Adapted Lesk) b- Context Vectors c- Hirst & St-Onge. Also define the WordNet as a machine readable dictionary created at the Cognitive Science Laboratory at Princeton University. Unlike most dictionaries, WordNet contains only open-class words (nouns, verbs, adjectives, and adverbs). WordNet does not contain closed-class words such as pronouns, conjunctions, and prepositions (Michelizzi J. 2005).

Rajan and lakshmi proposed incorporation of semantic matching methodology in semantic web for improving the efficiency and accuracy of the discovery mechanism. The current researches that aim

to developed Internet need dynamic retrieval and invocation methods. Also, authors claimed that better search capabilities are also rapidly emerging. But performance of keyword-based search engines cannot be termed efficient. In order to handle such problems, the Semantic technology can be incorporated using ontology and an efficient matchmaking methodology. Retrieval of relevant information can be done automatically, efficiently and accurately, thereby, reducing knowledge overhead and hence, manual data analysis. The number of searches and the time of search can also be vastly reduced. The current search is implemented using Vaccine Ontology in the Healthcare domain. This can be extended to the entire Healthcare domain. Also, other domains can acquire the capabilities of the web service discovery by using the same methodology, by using other ontology (Rajan J. and Lakshmi M. 2012).

Danger et al. addressed the problem of mapping text fragments into a given ontology in order to generate ontology instances that semantically describe this kind of resources. They can automatically populate a Semantic Web consisting of text documents that concern with a specific ontology. They have evaluated their approach over a real-application ontology and a text collection both in the Archeology domain (Danger R. et al. 2005).

CHAPTER 3

RE Ontology & Similarity Measures

3.1 Introduction

This chapter will discuss the preparation to build and use ontology in the domain of Renewable Energy. This preparation is done by collecting and analyzing many documents related to the field of RE applications. This chapter will discuss how to use these documents to design a model for extracting concepts in the domain of RE using KAON tool (Kayed A. et al. 2011). This research determine which semantic measure will be used in our work. Our approach aims to improve semantic matching by extracting instances that will help for this purpose.

3.2 Building Ontology

The process of developing an ontology in a specific domain is not easy and time consuming task (Kayed A. 2011). In order to do this task properly and to get the desired benefit from the ontology many parameters must be considered. These parameters aid in defining the final structure of ontology. The essential parameters we should identify at the beginning are the knowledge domain. The ontology will be covered and the intended users that will use it. The other parameter should be considered is specifying the motivation of ontology development; in other word identifying ontology purposes and usage. Identify all those parameters will be aid to determine the methodology for building the ontology.

Extract ontology concepts include many steps. First, collecting and studying large amount of documents which describe the domain. Second, extracted concepts from these documents using ontology building tools. Finally, studying, analyzing and filtering the extracted concepts in a way that serves the target of ontology around.

3.3 Extracting Renewable Energy Ontology Concepts

There are many ontology building' tools are available to assist in generate ontology candidate concepts and relationships. KAON is one of these tools, there are many reason to choose KAON tool such as:

1) It is open-source tool.
2) Easy creating and editing the ontology.
3) Building the ontology from underling textual data.
4) Easy for use.
5) The efficient user interface.

Maedche defined TextToOnto as: a tool suite built upon KAON in order to support the ontology engineering process by text mining techniques; providing a collection of independent tools for both automatic and semi-automatic ontology extraction (Maedche A. 2001). Also, Gabel stated KAON provides an API for programming management, providing access to ontologies using stand-alone server, moreover, its provided inference engine in order to solve the conjunctive queries (Gabel et al. 2004). See (Figure 3.1).

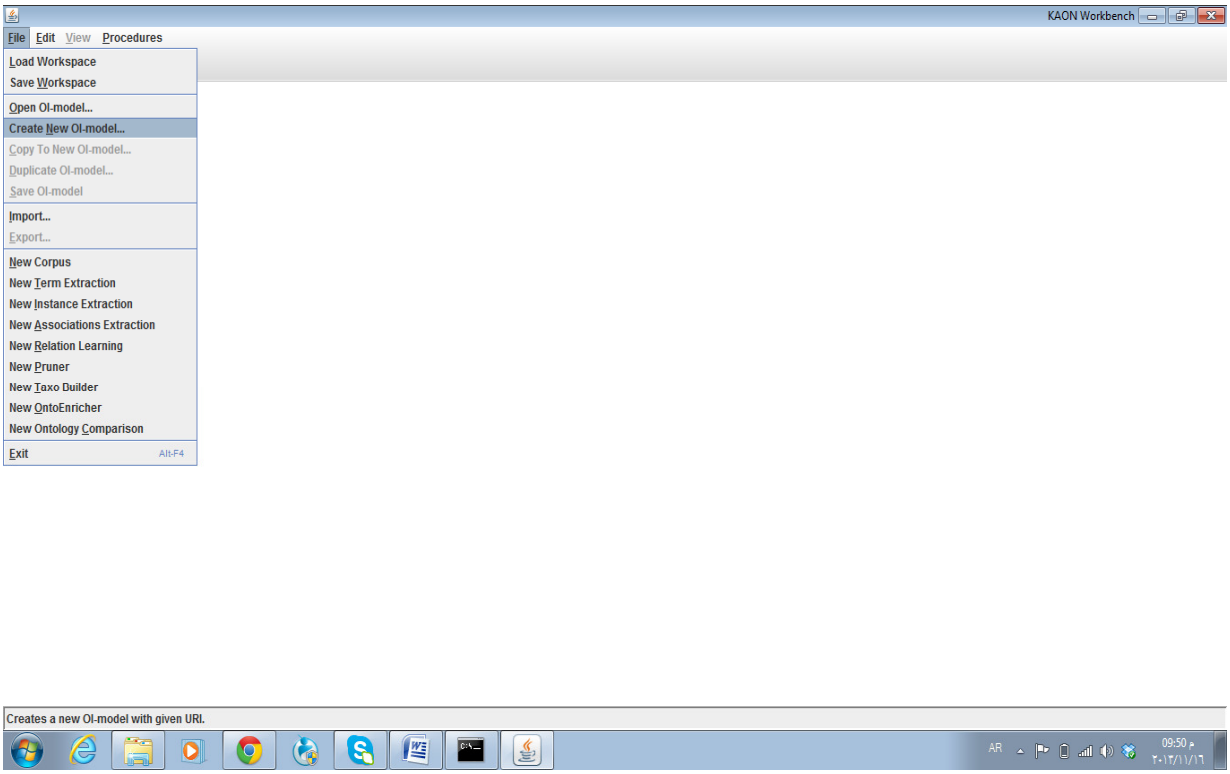


Figure 3.1: The front-end of the KAON TextToOnto Tool

To extract ontology concepts in the domain of renewable energy must follow several steps. First, we visited various RE providers, sites, documents, and reports and subscribe with them to collect their data. Second, converted each of documents into a text file. Finally, we added the prepared text corpus (from related documents) to the tool by using the new corpus function as showing in Figure 3.2.

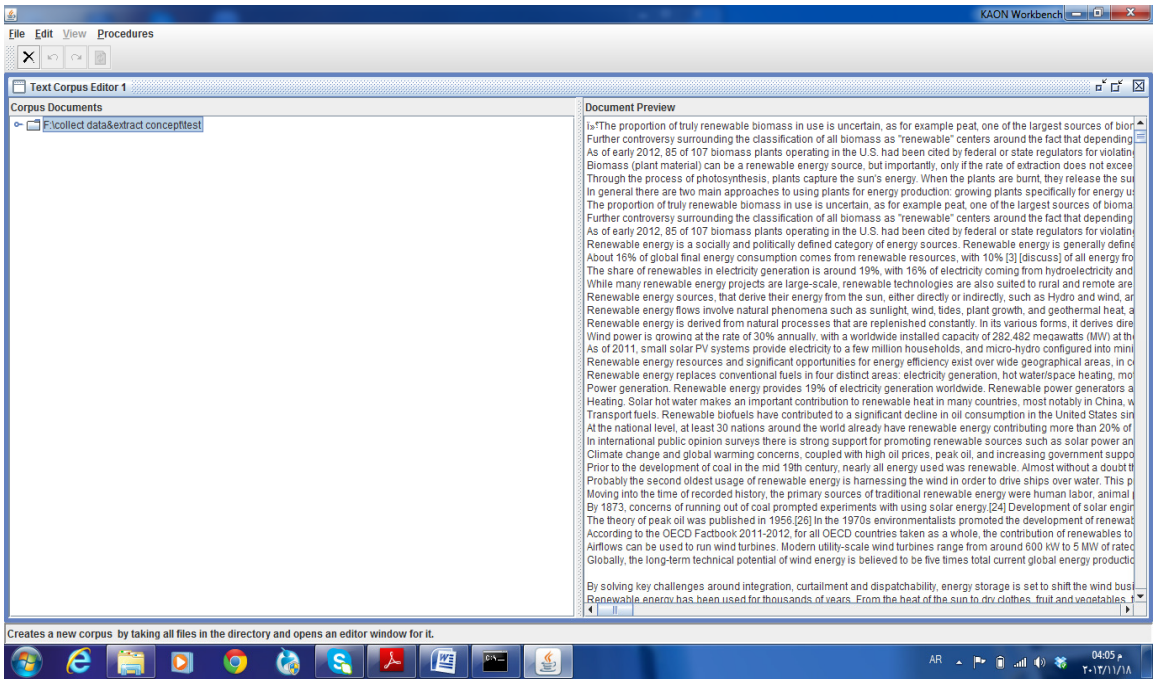


Figure 3.2: Create new corpus function using KAON TextToOnto Tools

Later we used the (New Term Extraction) function in order to extract concepts from the provided text corpus as showing in (Figure 3.3).

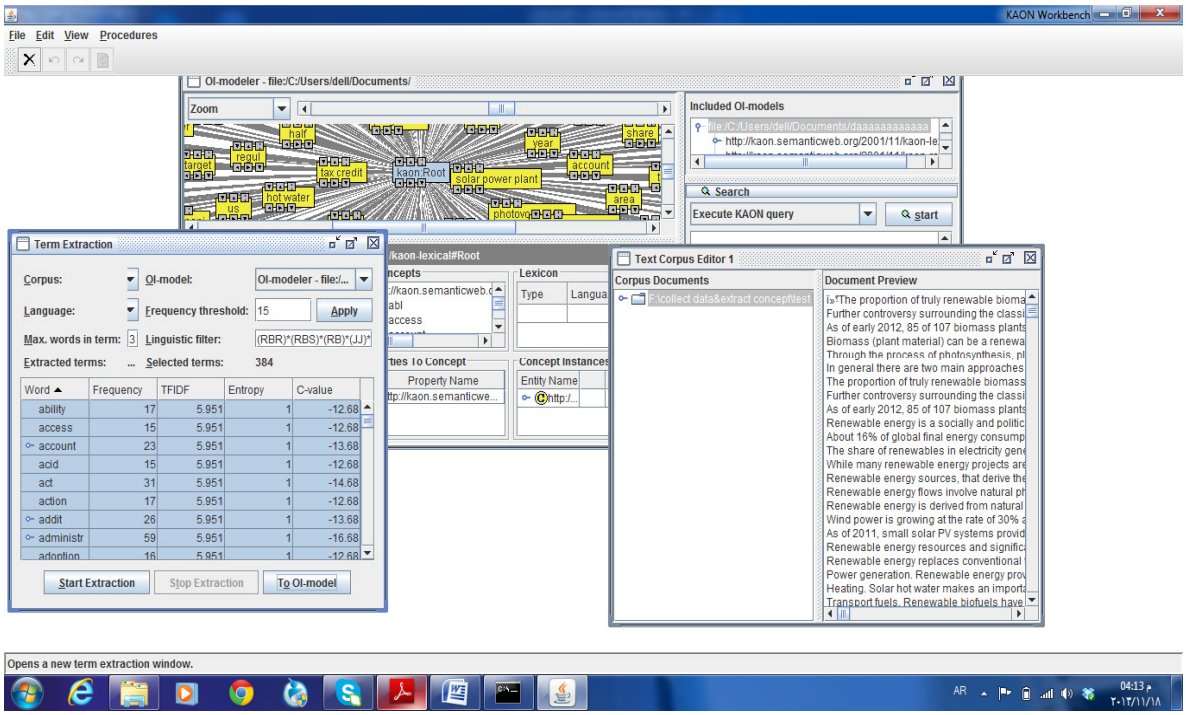


Figure 3.3: New Term Extraction function using KAON TextToOnto Tool

This tool extracts concepts using parameters; there are many frequency threshold available in the KAON ontology building's tool. First, we set the frequency parameter to 5 the result was 102 concepts, we present part of result in the table 3-1. For more details same table in the appendix. We refined the results of extraction ontology concepts by applying an elimination process for stopping words and characters.

Table 3-1: Results of extracting ontology concepts for RE

Frequency=5, No. of Concept=102

Air	Back	Billion
Biodiesel	Biomass	Capacity
Carbon	Cent	Change
Climate	Coal	Community
Consumption	Core	Cost
Demand	Develop	Earth
Efficient	Electricity	End
Form	Future	Gas

Then when we increased the frequency to (7), the result was 63 concepts. We present part of result in the table 3-2. For more details same table in the appendix.

Table 3-2: Results of extracting ontology concepts for RE

Frequency=7, No. of Concept=63

Billion	Biomass	Capacity
Carbon	Cent	Change
Climate	Coal	Cost
Consumption	Demand	Develop
Earth	End	Environment
Electricity	Fuel	Gas
Grid	Heat	Peat
Plant	Power	product

Then we increased the frequency threshold to 10, 13 and 15. The results were 38, 28 and 21 concepts respectively, for more details see tables (3-3, 3-4, and 3-5) in the appendix. We present the results of extracting concepts to experts in the field of RE, and they accepted the terms with frequency threshold (5). Because it has a set of important terms in the field (like: operation, support, temperature, ... etc). See table 3-3 in the appendix.

We repeated the preceding steps by adding more RE related documents, figure 3.4 present the ontology concepts of them. The extracted concept number differs depending on the collected document size.

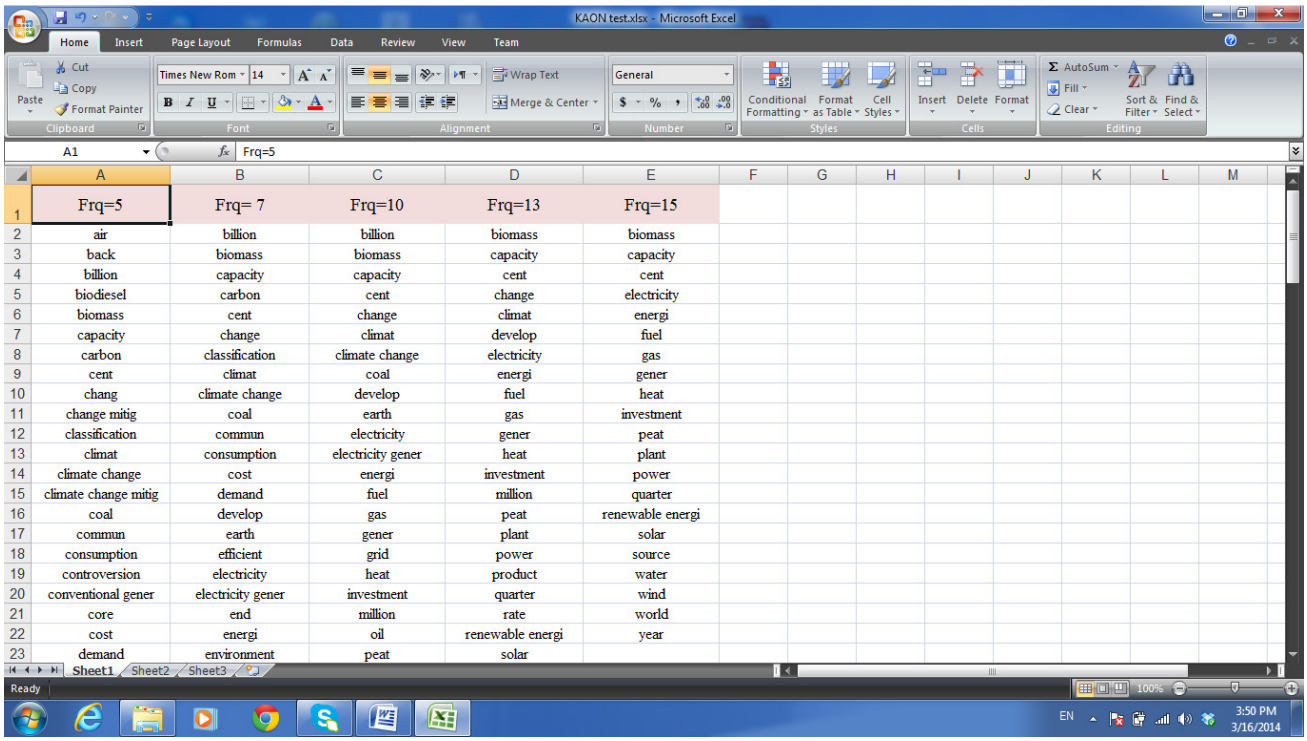


Figure 3.4: Concepts extraction for multi frequency

3.4 Semantic Matching Measures

The semantic similarity need a metric to find if the two concepts are related to each other or not, and if yes, how much? The key difference between the concepts should be considered in metrics, and not only the general shape, but the core meaning of the phrase will be taken into account. To determine if the two concepts are similar semantically, or not, a measurement of match should be computed. Michelizzi stated that there are many measurements are available for semantic matching (Michelizzi J. 2005). The following are some of the used measurements for similarity:

- Path Length : This measure computes the semantic relatedness of word senses by counting the number of nodes along the shortest path between the senses in the 'is-a' hierarchies of WordNet[1].
- Conceptual Distance : Describes the distance between different groups of society and is opposed to locational distance[2].
- Wu-Palmer : This measure calculates relatedness by considering the depths of the two synsets in the WordNet[1].
- Resnik : Defined the similarity between two synsets to be the information content of their lowest super-ordinate (most specific common subsumer)[1].

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

[2] http://www.sociologyguide.com/basic-concepts/Social-Distance.php

3 http://funsimmat.bioinf.mpi-inf.mpg.de/help3.php

- Lesk : Lesk (1985) proposed that the relatedness of two words is proportional to the extent of overlaps of their dictionary definitions. This LESK measure is based on adapted Lesk from Banerjee and Pedersen (2002) which uses WordNet as the dictionary for the word definitions. Computational cost is relatively high due to combinations of linked synsets to explore definitions, and need to process these texts[1].

3.4.1 Our Experiment

To determine the best measure for RE domain, we decided to test these measures according to a well-known standard (Resnik P. 1999). We apply the semantic similarity measures on the Resnik's standard sample. We apply these concepts on different measures such as (WUP, LIN, and Path) similarity measurements to determine which one will be used. We calculated the average for these concepts using the above similarity measures. These averages have been compared with the maximum for each measure. The results of this step show that WUP (similarity measure which is based on path lengths between the concepts to calculate the similarity between them). Table 3.6 present sample of standard concepts with different semantic similarity measures, for more details see table 3.6 in the appendix.

Table 3-3: Standard Concept Similarity Measures.

Concept 1	Concept 2	WUP	Error	LIN	Error	PATH	Error
Car	Automobile	100%	0%	100%	0%	100%	0%
Gem	Jewel	100%	0%	100%	0%	100%	0%
Journey	Voyage	95%	5%	82%	18%	50%	50%
. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .	. . . . .
tool	implement	94%	6%	91%	9%	50%	50%
Average			29%		54%		66%

This research will refer to Wu-Palmer method (WuP). Michelizzi illustrated this method by calculating the similarity between two concepts using the depth of the two concepts (Michelizzi J. 2005). Thus, LCS is defined by Baader as the Least Common Sub-summer. It represents the shared present between two nodes with deepest value and it is separated from the root concept. Where if the LCS goes deeper, in means that, the similarity measure is more large (Baader F. et al. 2004). As equation-2, also Baader F. stated the larger LCS gets larger Score in the similarity measure based on WUP approach for semantic measures. The LCS supports the knowledge based systems and makes the measurements more reasonable (Baader F. et al. 2004). Equation-2 shows how to calculate the WUP measure results, which is commonly known as score.

$$WUP(score) = 2 * \frac{depth\ (LCS)}{(depth(s1) + depth\ (s2))} \quad \text{.... (2)}$$

Where s1: is the first concept; s2: is the second concept.

(WUP) score : have the value larger than 0 and less than or equal to 1.

After the above experiment finished and refer to use WUP measure in our research. However, the range of WUP measure between 0 to 1. So, there are many cutting point threshold available in the WUP similarity measure. Thus, we need to determine which threshold use in this research. Our experiments starting with medium score value equal 0.5, it could be clear that this is not good threshold value and it should be greater. Then chosen a 0.6, 0.7 and 0.8 value, then 0.9. The methodology that is being used to determine the best threshold in this research is simple and based on trial and error. A set of instances where selected and the WUP score is calculated continuously for them. Then, by human experience, it is easy to determine the best value that could be useful in such application. The deal with larger value that is near to 1 is not suitable. Hence, it will make the matching pass only for very similar instances, where those may be the same (values near to 1 is like 0.97). So, in this research after the process of trial and error, a 0.9 score is considered to be used. The complete program is designed over this value, and the recorded results were taken based on it. Actually, trying to use value larger makes the matching process more constrained, and the matching process will not pass successfully. Also, chosen a lower value for the score threshold will cause the decision making process to be unreasonable, while selecting many instances to be similar, but actually they are not. The details of the above experiments to determine the best threshold will be discussed in the chapter five.

Chapter 4

Methodology and Proposed Model

4.1 Overview

This chapter explains in details, the proposed model and its components. This will be done through the design and implementation of extracting instances to improve semantic matching. In this chapter, we will discuss the instances-based data extraction (IDE) algorithm for extracting instances and explain how we modified this algorithm. Section two show the proposed system. Section three shows how the semantic matching will be enhanced using instances. This model will measure how much a renewable energy is close (semantically) from consumers' requirements.

4.2 The Original Instances-based Data Extraction algorithm

Murthy and Reddy proposed an algorithm named instances-based data extraction (IDE) for extracting instances (Murthy K. & Reddy V. 2013). This algorithm works as follows:

1- Selected a page randomly.
2- This page is given to the user for labeling by prefix and suffix.
3- The system starts to extract items from new pages. Compares the stored prefix and suffix with the tag stream of new page to extract each item.
4- If some items from new pages cannot be identified, then this page is passed to the user for labeling.
5- This process goes on until all the pages from the given site have been processed.

The above algorithm does not has the learning step and thus saves the learning time and manual labeling is implemented which is labor intensive and time consuming (Murthy K. & Reddy V. 2013). Also, this algorithm use the keywording matching. Additionally, the process of using and choosing prefix and suffix for each mark item is very difficult. Therefore, all these reasons lead to avoiding using this algorithm.

4.2.1 Modify IDE Algorithm for Extracting Instances

This research proposes an enhanced algorithm that merges instances extraction and semantic matching in application of renewable energy. Thus, we modified the IDE algorithm that will help in the extracting instances process. The algorithm works after modified as follows:

1- Collected huge data related to the RE domain.
2- Extracting instances manually from a page selected randomly.
3- Assign one to the item counter (frequency) for each extracted item and put them in a list.
4- For new page, compares each item in the list with all items in the new page.
5- If there is a match the items counter will be increased by one.
6- Else, compares the item with the instances list using WUP similarity measure.
7- (IF WUP >= a threshold) then extract item and updated the list of instances.
8- This process goes on until all pages have been processed.

The main idea of IDE enhancment algorithm is to extract a set of instances that are close in meaning to the marked objects. Richard stated there are many algorithm could be adopted and used for instance detection. There is no exact expression of instances in different applications and object properties (Richard C. and William W. 2009).

The modified algorithm is a method for extracting instances which performs extraction by comparing each new instance or page in order to be extracted with a labeled instance from previous step. For a new page, the algorithm compares the stored items with the items from a new page to extract each item. If there is a match with one of them this will increase their frequency and will be extracting these words. Otherwise, if some items from the new page cannot be identified, a comparison will occur with the same list using WUP matching measure. If larger or equal specific threshold then it will be extracted. Else the item will be cancelled. For example, the word (wind) isn't found when we compared with the list of items, so the comparison process using WUP shown (wind WUP air >= threshold) true. That means this word will be updated on the list manually with new item. This process continues until all items are extracted or the pages are completed. The next step is implement the above algorithm for extracting instances in the domain of RE.

4.2.2 Applying the Modified Algorithm

After modified the (IDE) algorithm, the next step is implement the algorithm to extract instances in our domain. We have done the following:

1- Collect huge data related to the RE domain from web sites.

2- The user initially open a random renewable energy related page.

3- Then select each item inside the page and mark it manually.

4- Apply the algorithm on the rest pages.

The marking process will be done by the user. In the random page, usually the first page, the user selects several items which represent the items of interest in the renewable energy field. Then the user has to save the extracted instances manually in a separated text files that represents the marked page.

The marked item will be saved with their frequencies. The file will be used by the application to find all items that considered as instances with their frequency. Table 4-1 shows a sample of the marked instances that is being extracted for the renewable energy manually. The first column of this table represents the marked items, where the second column shows the frequency of that item in the selected page.

Table 4-1: Sample of Extracting Instances Manually

Jordan	1
IT	1
Price	3
Offer	2
Air	4
H2O	1
Yr	1
Voltage	3
Sunshine	1

Documents have been collected from the web in the domain of RE. the collection pried has been spanned from 5/10/2013 to 1/11/2013. The original size of the document was 3 MB. All documents have been converted to text file and feed them to our application with starting instances of 23. These are the manual extracted instances. After we run the application, the number of instances have been 72. The following table give a sample of these instances with frequency.

No.	Instances	Freq.
1	Power	84
2	Jordan	1
3	IT	4
4	Price	1
5	Content	1
6	Home	7
7	Offer	6
8	Air	3
9	Range	12
10	Month	2
11	Solarize	2
12	Mean	1
13	Water	30
14	Capacity	11
15	Addition	2

Table 4-2: sample of extracting instances

4-3 Tested the Proposed Work

Nine providers of renewable energy have been selected according to (company's capital, experiences, available services,…, etc). After selecting the providers, we wrote 30 queries that represents the customer requirements. The application has been using Visual Studio in Visual Basic. The implemented program enables the customer to directly fills his / here requirements query in a ".txt" format file. The program will read it directly from the file path. The algorithm will run on each the provider's file and on each query (30 queries) . Figure 4-1 shows the user interface of the application.

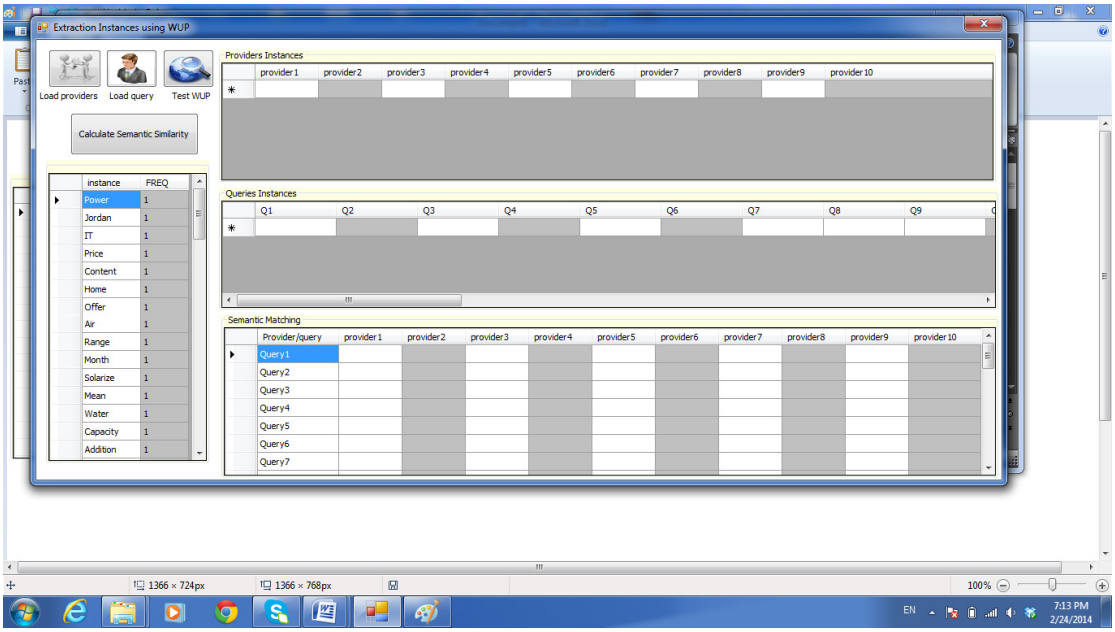


Figure 4-1: User Interface for the application

This figure contains list of instances that extracted manually by human. Load provider key for loading the providers text files to extract the instances from them. Also, load query key for loading the queries text files to extract the instances from them. Additionally, semantic similarity measure key for measuring the similarity between above two table of instances (providers instances, query instances). We will discuss the application in details in the next section.

4.3.1 Extracting Instances from Providers and Queries

Now assuming that the selected item from provider or query files does not exist in the manually marked items and not similar to any one of the marked item then we apply enhanced algorithm. The extracted instances will be either existing in the manually marked items file or semantic matching will be implemented in order to confirm if it matches the WUP policy.

When instances are extracted, they are saved in a table, that is arranged based on the different provider profiles. Each column in the table contains the instances that are related to the providers. Then instances could be extracted directly from provider profile and saved in table to use it later. The instances which are extracted from the queries' files, will be handled in the same way as the provider files. The same algorithm for instances extraction is used. The same illustration figures describe the query instances extraction with the change of the input text file only. All these instances will be stored and matched. In other word, all instances for each queries will be matched with all instances of each 9 providers.

After extracting the matched instances, one more process is to be handled in order to take a complete decision, that is which of the providers is better to supply the renewable energy. The percentage of matches to unmatched will be calculated by dividing the total number of instances for each query to the number of matched instances. The provider with maximum percentage will be considered the best provider. This means that provider meets the customer's requirements in renewable energy field.

This process will help to decide which provider is more suitable for specific query. The fact is that application target of this presented system is to help the customer to decide which provider is more appropriate to supply the renewable energy to customers. By this algorithm, the instances have improve semantic matching as illustrated in problem statement in chapter-1.

This research concerns to implement semantic matching between two types of extracted instances; the instances that are extracted from the customer queries and the instances that are extracted instances from the providers text file. Thus, this research implements semantic matching algorithm based on (WUP) semantic measure algorithm.

Souleiman illustrated that the adoption of semantic matching will solve all problems those related to text semantics, including the meaning semantics and the lexical structure semantics. It will give rigid comparison criteria to find the similarity (Souleiman H. et al. 2012).

In this research the process of semantic matching is conducted in two methods. First, matching between the structure table of the instances that are extracted from the providers profiles and the structure table of the instances that are extracted from the queries text files without using the frequency, for example:

Query Instances1: company, power, energy, wind, design, operations and years.

Provider Instances1: generation, design, performance, wind, operations, production, capabilities, capacity, power, current, technology, company, years, installation, quality.

We have 7 instances from query1 and 15 instances from provider1. Now the semantic matching between them shows 6 matched items from 7 items that means that the average of matching between QI 1 with PI 1 (6 / 7) = 86%.

	PI1	PI2	PI3	……...					PI9
QI1	89%	67%	33%						67%
QI2	50%	63%	38%						50%
QI3	36%	55%	18%						45%
QI4	43%	71%	43%						29%
QI5	27%	45%	45%						45%
:									
:									
:									
:									
:									
QI30	44%	78%	44%						89%

Table 4-3: Sample of semantic matching.

Then, these customer queries and RE providers information are sent to human experts that participant in the field of renewable energy. However, we asked them to help by giving us how much each customer query achieve with RE providers. After a while the results were sent back from experts. It were collected and studied upon the results of all experts. They all evaluated as a good example for the studied field services and they give us in percentage how much each provider achieve from the customer's requirements. Table 4-4 presents sample of the result of this step. For more details see table 6 in the appendix.

	PI1	PI2	PI3	…….					PI9
OI1	70%	60%	55%						60%
OI2	60%	70%	50%						50%
OI3	50%	65%	40%						50%
⋮									
⋮									
⋮									
⋮									
⋮									
⋮									
OI30	50%	70%	60%						75%

Table 4-4: Sample of expert human results

The error was calculated by subtracting human percentage from the matching percentage. Equation 1, 2 was used to calculate the percentage of error.

$$Ps = \frac{\sum(the\ Number\ of\ Match\ Concepts)*100}{\sum Average\ Weight}\ \% \qquad \ldots 1$$

Ps is the provider percentage.

Then

$$error = |\ Ps - H\ | \qquad \ldots 2$$

H is the human expert percentage

The matching error is calculated for each query with all providers in the same way to know which provider retrieves the highest percentage.

	(P)1			**……**					**(P)n**
(Q)1	SM	H	E						
:									
:									
:									
:									
:									
:									
:									
:									
:									
:									
:									
(Q)m									

Second method is semantic matching when we using the frequency. Thus, some of the concepts/instances have high frequency which may affect on the results, for that we take the average. We calculate the average of frequencies then normalize the frequency for the instances list that are obtained through equation-3. Then choose the minimum number between (1, Frequency Normalization) through using equation-4.

Freq. per Avg. = Frequency of Item / Average of Frequency …(3)

Normalized Freq. = MIN (1, Freq. per Avg.) …(4)

Afterwards, semantic matching between the structure table of the instances is extracted from the providers profiles, and the structure table of the instances that are extracted from the queries text files depending on the equation-4. The error was calculated by subtracting human percentage from the matching percentage. Equation 1, 2 was used to calculate the percentage of error. We will present the results for all these processes in details in the next chapter.

4.4 Decision Making

Kenji defined that the decision making process is the process of getting some of providers' profiles, and inputting customers' specific queries (Kenji D. and Michael N. 2012). Moreover, applying the enhancement algorithm of this research in order to determine which provider is the most match to the customer queries to supply it with renewable energy service. The percentage of match is calculated depending on how a specific provider scores and how many matches it achieves with all queries. In other words, the percentage of match is a function of semantically matched instances and the frequency of matched instances. Miller stated phrase matching is required in many applications, like extraction of information, retrieval of information, automatic translation, copyright infringement, automatic identification processes and other applications (Miller J. 2011). In this research the application of phrases matching is to make a decision in the field of renewable energy regarding which provider is suitable for specific customer query. This decision is related to determining which provider from a list of well known providers are the best match for specific customer input query. The next step is to compare the match results and determine the maximum value of matching and the provider that achieves that value. The decision making will be a selection of the provider that supplies better renewable energy options, services and other features. The best match provider depends on the customer queries and the ability of the user to determine the exact queries for the renewable energy and the case of use. It also depends on the provider itself and the ability to supply

specific renewable energy features or services to the customers. In fact, if the user queries are not correct as possible as enough to get the scope of his / here required application, the match percentage will be not complete depending on his query. So, it is important to select the correct reasonable query from the customer. This differs from one case to another and depends on the case of providers, consumer specs, location and the cost of that supplied system in the determined case. Figure 4-2 shows the flow chart of the match percentage calculation and the decision making process.

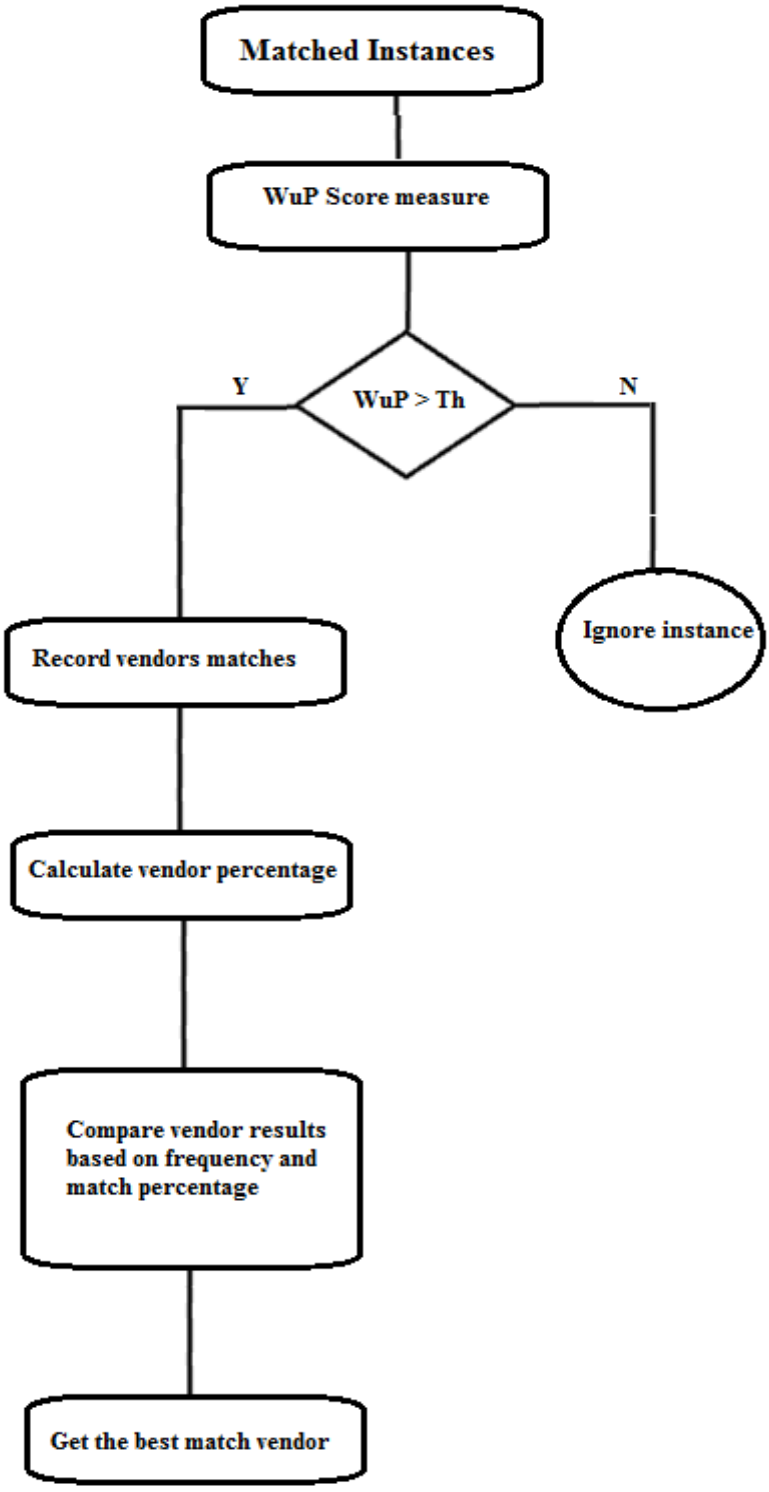


Figure 4-2: Flow chart of match percentage calculation and decision making.

Finally, figure 4-3 illustrates the total block diagram of the presented system in this research with all components, starting by customer input queries and supplied provider files. Also includes the instances extraction process from both, the provider files and the input queries. However, it includes the semantic matching and the semantically resulted instances, in addition to the decision making process after measuring the best match provider based on the match percentage.

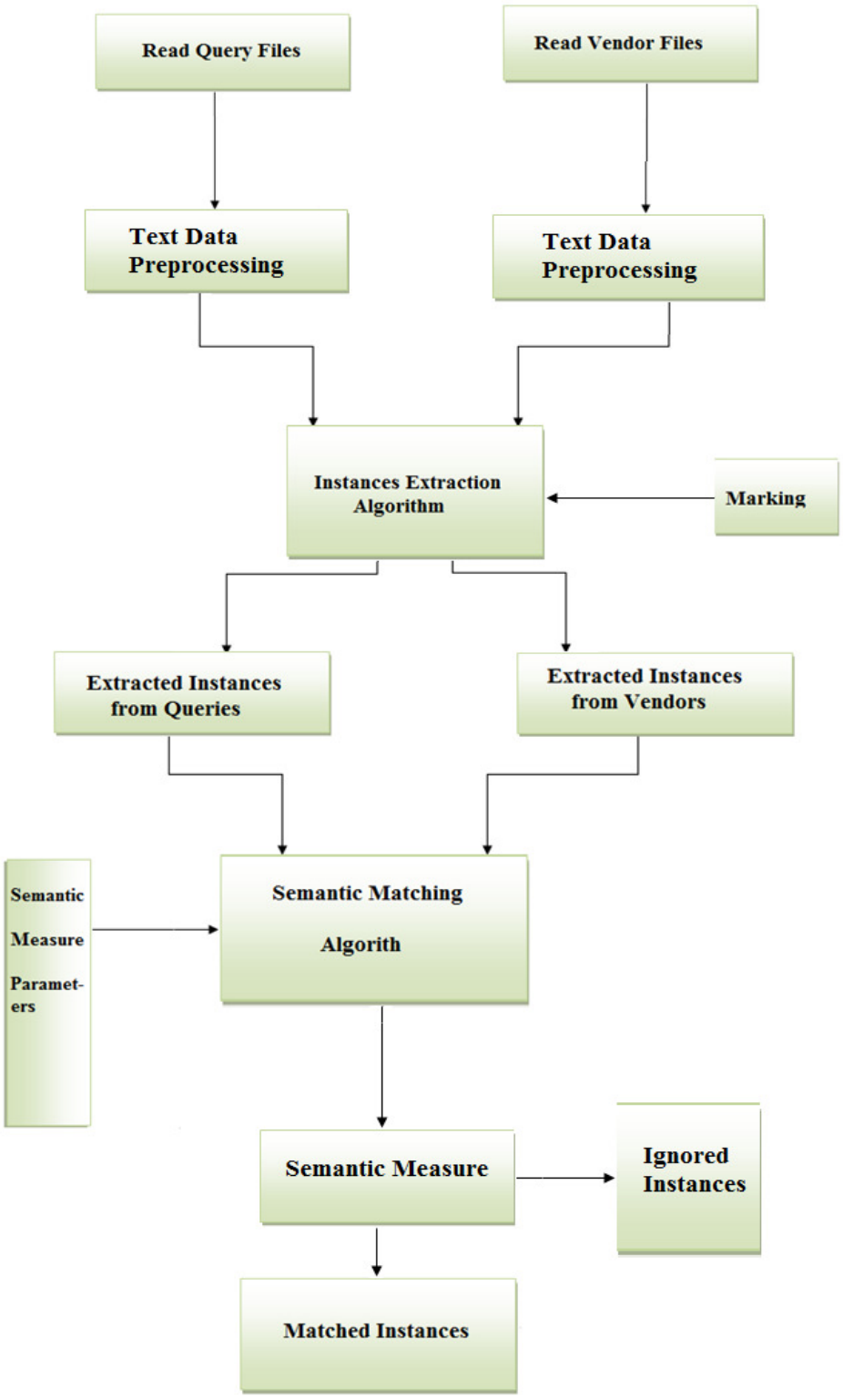


Figure 4-3: Main block diagram of the presented thesis

CHAPTER 5

Experimental Results

5.1 Overview

The application that was designed in order to implement the algorithm that is presented in this thesis aims to extract the instances that were used in the semantic matching of the proposed and presented algorithm. In addition, it is important to record the results of the presented algorithm. Thus, the experiments in section 5.2, 5.3 depending on the used method in the previous chapter. Analysis of the results will be discussed in section 5.4. Finally, section 5.5 discusses main contributions in this research.

5.2 Semantic Matching Methods

Two types of results for the semantic matching according to the used method. The results will be discussed in details in the next section. The experiment has been repeat (1620) times with different parameters. First we compute the percentage of matching for:

1) The concepts only.
2) The concepts and instances.
3) The instances only, we used five threshold for WUP (0.5, 0.6, 0.7, 0.8, 0.9).

In all cases we used 9 providers with 30 queries. The total number of experiments was:

Total no. of Exp.= [no. of process * no. of provider * no. of query]

= [6 * 9 * 30]

= 1620 times

After that the error has been computed according to the human experts when we assume the experts is the standard and the error has been computed by take the different absolute error of each reading. The average of error has been computed before we go in details. The following table summarized the results:

Approach	WUP >= 0.9	WUP >= 0.8	WUP >= 0.7	WUP >= 0.6	WUP >= 0.5	KAON concepts
Error without Frequency	10%	26%	34%	35%	36%	27%

Table 5-1: average error for semantic matching

The above table shows that the average error is minimal when we use instances and WUP measure with threshold of (0.9). It also indicates that if the WUP threshold is less than (0.9) using only the concepts (column 6) will be almost the some as using instances. The average error will be around (27%). In the following we will discuss:

1) Concepts only.
2) Instances with WUP >= 0.5.
3) Instances with WUP >= 0.6.
4) Instances with WUP >= 0.7.
5) Instances with WUP >= 0.8.
6) Instances with WUP >= 0.9.

5.2.1 Semantic Matching with KAON concepts

We used the traditional keyword matching technique to match each query with all providers. This technique matches the query instances with providers instances. The results presents percentage the numbers of instances shared between it. Table 5.2 presents sample of the results of matching according to used concepts only. For more detail see table 17 in appendix.

	P1	Human	**Error**	P2	Human	**Error**
Q1	100%	70%	**30%**	75%	60%	**15%**
Q2	50%	60%	**10%**	50%	70%	**20%**
Q3	50%	50%	**0%**	50%	65%	**15%**
Q4	0%	55%	**55%**	0%	60%	**60%**
Q5	50%	40%	**10%**	25%	55%	**30%**
Q6	25%	50%	**25%**	50%	50%	**0%**
Q7	67%	45%	**22%**	67%	75%	**8%**
Q8	50%	40%	**10%**	50%	60%	**10%**
Q9	50%	50%	**0%**	0%	65%	**65%**
Q10	50%	65%	**15%**	50%	60%	**10%**

Table 5-2: Presents sample of the results of semantic matching with KAON concepts

If we take query 5 in the above table, provider1 presents 10% as an error. This error presents when subtract the percentage of human expert (40%) with the percentage of (Q1 VS P1) (50%). The overall error using this technique was 27%.

5.2.2 Semantic Matching WUP >= 0.5

Table 5.3 presents sample of the results of matching according to used instances and WUP threshold less than or equal 0.5. The overall error using this technique was 36%. For more detail see table 15 in appendix.

	Provider1	Human	**Error**	Provider2	Human	**Error**
Q 1	32%	70%	**38%**	36%	60%	**24%**
Q 2	22%	60%	**38%**	30%	70%	**40%**
Q 3	24%	50%	**26%**	29%	65%	**36%**
Q 4	19%	55%	**36%**	31%	60%	**29%**
Q 5	18%	40%	**22%**	35%	55%	**20%**
Q 6	18%	50%	**32%**	41%	50%	**9%**
Q 7	25%	45%	**20%**	38%	75%	**37%**
Q 8	29%	40%	**11%**	43%	60%	**17%**
Q 9	29%	50%	**21%**	29%	65%	**36%**
Q 10	26%	65%	**39%**	37%	60%	**23%**

Table 5-3: Presents sample of the results of semantic matching with WUP >= 0.5.

5.2.3 Semantic Matching WUP >= 0.6

Table 5.4 presents sample of the results of matching according to using instances and WUP measure with threshold less than or equal 0.6. The overall error using this technique was 35%. For more detail see table 12 in appendix.

	Provider1	Human	**Error**	Provider2	Human	**Error**
Q 1	38%	70%	**32%**	33%	60%	**27%**
Q 2	22%	60%	**38%**	30%	70%	**40%**
Q 3	24%	50%	**26%**	29%	65%	**36%**
Q 4	20%	55%	**35%**	33%	60%	**27%**
Q 5	19%	40%	**21%**	38%	55%	**17%**
Q 6	18%	50%	**32%**	41%	50%	**9%**
Q 7	24%	45%	**21%**	35%	75%	**40%**

Q 8	29%	40%	**11%**	43%	60%	**17%**
Q 9	29%	50%	**21%**	29%	65%	**36%**
Q 10	26%	65%	**39%**	37%	60%	**23%**

Table 5-4: Presents sample of the results of semantic matching with WUP >= 0.6.

5.2.4 Semantic Matching WUP >= 0.7

Table 5.5 presents sample of the results of matching when we using instances and WUP measure with threshold less than or equal 0.7. The overall error using this technique was 34%. For more detail see table 9 in appendix.

	Provider1	Human	**Error**	Provider2	Human	**Error**
Q 1	40%	70%	**30%**	35%	60%	**25%**
Q 2	24%	60%	**36%**	33%	70%	**37%**
Q 3	28%	50%	**22%**	33%	65%	**32%**
Q 4	20%	55%	**35%**	33%	60%	**27%**
Q 5	19%	40%	**21%**	38%	55%	**17%**
Q 6	18%	50%	**32%**	41%	50%	**9%**
Q 7	25%	45%	**20%**	38%	75%	**37%**
Q 8	29%	40%	**11%**	43%	60%	**17%**
Q 9	29%	50%	**21%**	29%	65%	**36%**
Q 10	26%	65%	**39%**	37%	60%	**23%**

Table 5-5: Presents sample of the results of semantic matching with WUP >= 0.7.

5.2.5 Semantic Matching WUP >= 0.8

Semantic matching technique was used in this experiment to match the query instances and providers instances. Table 5-6 presents sample of the results of matching according to using instances and WUP measure with threshold 0.8. For more detail see table 6 in appendix.

	Provider1	Human	**Error**	Provider2	Human	**Error**
Q 1	47%	70%	**23%**	41%	60%	**19%**
Q 2	31%	60%	**29%**	44%	70%	**26%**
Q 3	33%	50%	**17%**	50%	65%	**15%**
Q 4	27%	55%	**28%**	45%	60%	**15%**
Q 5	21%	40%	**19%**	43%	55%	**12%**
Q 6	20%	50%	**30%**	47%	50%	**3%**
Q 7	33%	45%	**12%**	50%	75%	**25%**
Q 8	44%	40%	**4%**	67%	60%	**7%**
Q 9	33%	50%	**17%**	33%	65%	**32%**

Table 5-6: Presents sample of the results of semantic matching WUP >= 0.8.

If we take query 5 in table 5.6 as an example, provider1 presents (21%) as a result of the semantic matching. This query presents 19% as an error. This error presents when subtract the human expert percentage (40%) from the semantic matching between (Q5 VS P1) (21%). The overall error using this technique was 26%.

5.2.6 Semantic Matching WUP >= 0.9

We used semantic similarity measure technique WUP >= 0.9 to match customer requirement with the appropriate provider. In order to do that, this technique matches the query instances with providers instances. The results presents percentage the numbers of instances shared between it. Table 5.7 presents sample of the results of matching according to using instances and WUP measure with threshold less than or equal 0.9. For more detail see table 3 in appendix.

	P1	Human	**Error**	P2	Human	**Error**
Q1	89%	70%	**19%**	67%	60%	**7%**
Q2	50%	60%	**10%**	63%	70%	**7%**
Q3	36%	50%	**14%**	55%	65%	**10%**
Q4	43%	55%	**12%**	71%	60%	**11%**

Q5	27%	40%	**13%**	45%	55%	**10%**
Q6	27%	50%	**23%**	64%	50%	**14%**
Q7	57%	45%	**12%**	86%	75%	**11%**
Q8	50%	40%	**10%**	75%	60%	**15%**
Q9	57%	50%	**7%**	57%	65%	**8%**
Q10	50%	65%	**15%**	70%	60%	**10%**

Table 5-7: Presents sample of the results of semantic matching WUP >= 0.9.

If we take query 5 in the above table, provider1 presents 13% as an error. This error presents when subtract the percentage of human expert (40%) with the percentage of (Q1 VS P1) (27%). The overall error using this technique was 10%.

5.3 Semantic Matching after Frequency Normalized

The previous experiments showed the results without giving any attention to frequency. This section show the affect of the frequency on the results. All 6 process with 9 providers and 30 queries have been repeated to take the frequency in consideration. The frequency has been record for all concepts and instances in the building process table.

Approach	KAON concepts only	Instances with WUP >= 0.5	Instances with WUP >= 0.6	Instances with WUP >= 0.7	Instances with WUP >= 0.8	Instances with WUP >= 0.9
Error with Frequency	25%	22%	21%	21%	17%	10%

Error without Frequency	27%	36%	35%	34%	26%	10%

Table 5-8: presents the overall error for all experiments.

Table 5-8 showed the difference between results in the two methods of experiments. Also it showed the affect of frequency when we use it in the experiments. Thus, using of frequency has been improved the results of matching. However, some of the concepts/instances have high frequency which may affect on the results, for that we take the average through:

Freq. per Avg. = Frequency of Item / Average of Frequency.

Now, the item that have high frequency can get the maximum average (1). Else, choose the minimum number between (1, Frequency Normalization) through using the following equation :

Normalized Freq. = MIN (1, Freq. per Avg.).

Table 5-9 presents sample of results for the frequency per average.

Instances	Freq.	Norm. Freq.
Power	84	1
Jordan	1	0.17
IT	4	0.67
Price	1	0.17
Content	1	0.17
Home	7	1
Offer	6	1
Air	3	0.5
Range	12	1
Month	2	0.33

Solarize	2	0.33
Mean	1	0.17
Water	30	1
Capacity	11	1
Addition	2	0.33
Design	18	1
Vigor	1	0.17
Iraq	1	0.17
Engineering	25	1

The above method shows that the average error is minimal when we use instances and WUP measure with threshold of (0.9). It also indicates that if the WUP threshold is less than (0.9) using only the concepts (column 6) will be almost the same as using instances. The average error will be around (25%). In the following we will discuss:

1) Concepts only.
2) Instances with WUP >= 0.5.
3) Instances with WUP >= 0.6.
4) Instances with WUP >= 0.7.
5) Instances with WUP >= 0.8.
6) Instances with WUP >= 0.9.

5.3.1 Semantic Matching with KAON concepts

Table 5.10 presents sample of the results of matching according to this approach. The overall error using the concepts only was 25%. For more detail see table 23 in appendix.

	P1	Human	Error	P2	Human	Error

Q1	100%	70%	30%	75%	60%	15%
Q2	50%	60%	10%	50%	70%	20%
Q3	50%	50%	0%	50%	65%	15%
Q4	0%	55%	55%	0%	60%	60%
Q5	50%	40%	10%	25%	55%	30%
Q6	25%	50%	25%	50%	50%	0%
Q7	67%	45%	22%	67%	75%	8%
Q8	50%	40%	10%	50%	60%	10%
Q9	50%	50%	0%	0%	65%	65%
Q10	50%	65%	15%	50%	60%	10%

Table 5-10: Presents sample of the results of semantic matching with KAON concepts.

5.3.2 Semantic Matching WUP >= 0.5

Table 5.11 presents sample of the results of matching according to using instances and WUP measure with threshold less than or equal 0.5. The overall error using this technique was 22%. For more detail see table 22 in appendix.

	P1	Human	Error	P2	Human	Error
Q1	54%	70%	16%	42%	60%	18%
Q2	33%	60%	27%	41%	70%	29%
Q3	31%	50%	19%	39%	65%	26%
Q4	35%	55%	20%	58%	60%	2%
Q5	24%	40%	16%	35%	55%	20%
Q6	24%	50%	26%	55%	50%	5%
Q7	43%	45%	2%	54%	75%	21%
Q8	30%	40%	10%	60%	60%	0%
Q9	46%	50%	4%	46%	65%	19%
Q10	36%	65%	29%	43%	60%	17%

Table 5-11: Presents sample of the results of semantic matching with WUP >= 0.5.

5.3.3 Semantic Matching WUP >= 0.6

Table 5.12 presents sample of the results of matching according to using instances and WUP measure with threshold less than or equal 0.6. The overall error using this technique was 21%. For more detail see table 21 in appendix.

	P1	Human	Error	P2	Human	Error
Q1	59%	70%	11%	46%	60%	14%
Q2	33%	60%	27%	41%	70%	29%
Q3	31%	50%	19%	39%	65%	26%
Q4	36%	55%	19%	60%	60%	0%
Q5	24%	40%	16%	35%	55%	20%
Q6	24%	50%	26%	57%	50%	7%
Q7	43%	45%	2%	54%	75%	21%
Q8	32%	40%	8%	64%	60%	4%
Q9	46%	50%	4%	46%	65%	19%
Q10	36%	65%	29%	43%	60%	17%

Table 5-12: Presents sample of the results of semantic matching with WUP >= 0.6.

If we take query 5 in the above table, provider1 presents 16% as an error. This error presents when subtract the percentage of human expert (40%) with the percentage of (Q1 VS P1) (24%).

5.3.4 Semantic Matching WUP >= 0.7

Table 5.13 presents sample of the results of matching according to using instances and WUP measure with threshold less than or equal 0.7. The overall error using this technique was 21%. For more detail see table 20 in appendix.

	P1	Human	Error	P2	Human	Error
Q1	60%	70%	10%	48%	60%	12%
Q2	35%	60%	25%	43%	70%	27%
Q3	33%	50%	17%	42%	65%	23%
Q4	36%	55%	19%	60%	60%	0%

Q5	24%	40%	16%	35%	55%	20%
Q6	24%	50%	26%	57%	50%	7%
Q7	45%	45%	0%	56%	75%	19%
Q8	32%	40%	8%	64%	60%	4%
Q9	46%	50%	4%	46%	65%	19%
Q10	36%	65%	29%	43%	60%	17%

Table 5-13: Presents sample of the results of semantic matching with WUP >= 0.7.

5.3.5 Semantic Matching WUP >= 0.8

Table 5.14 presents sample of the results of matching according to using instances and WUP measure with threshold less than or equal 0.8. The overall error using this technique was 17%. For more detail see table 19 in appendix.

	P1	Human	Error	P2	Human	Error
Q1	65%	70%	5%	51%	60%	9%
Q2	40%	60%	20%	49%	70%	21%
Q3	40%	50%	10%	50%	65%	15%
Q4	43%	55%	12%	72%	60%	12%
Q5	26%	40%	14%	37%	55%	18%
Q6	26%	50%	24%	60%	50%	10%
Q7	52%	45%	7%	65%	75%	10%
Q8	39%	40%	1%	78%	60%	18%
Q9	50%	50%	0%	50%	65%	15%
Q10	38%	65%	27%	46%	60%	14%

Table 5-14: Presents sample of the results of semantic matching with WUP >= 0.8.

5.3.6 Semantic Matching WUP >= 0.9

Semantic matching technique was used in this experiment to match the query instances and providers instances. Table 5-15 presents sample of the results of matching according to using instances and WUP measure with threshold less than or equal 0.9. For more detail see table 18 in appendix.

	P1	Human	Error	P2	Human	Error
Q1	**89%**	70%	**19%**	**68%**	60%	**8%**
Q2	**43%**	60%	**17%**	**72%**	70%	**2%**
Q3	**49%**	50%	**1%**	**61%**	65%	**4%**
Q4	**56%**	55%	**1%**	**94%**	60%	**34%**
Q5	**43%**	40%	**3%**	**43%**	55%	**12%**
Q6	**30%**	50%	**20%**	**70%**	50%	**20%**
Q7	**73%**	45%	**28%**	**91%**	75%	**16%**
Q8	**42%**	40%	**2%**	**70%**	60%	**10%**
Q9	**65%**	50%	**15%**	**65%**	65%	**0%**
Q10	**51%**	65%	**14%**	**66%**	60%	**6%**

Table 5-15: Presents sample of the results of semantic matching with WUP >= 0.9.

If we take query 2 in the above table as an example, provider2 presents 72% as a result of the semantic matching. The query presents 2% as an error. This error presents when subtract the semantic matching (72%) from the human expert percentage (70%). The overall error using this technique was 10%.

5.4 Analysis of the Results

This section summarizes and analyzes the experiments results conducted in this chapter. The experiments results showed that the error rate was decreases gradually when we using different cutting points for WUP. Additionally, the use of frequency and how can affect on the results. Two type of results have been showed depending to using frequency or not. The experiment from the first method without using frequency showed (Semantic Matching using concepts only) results in 73% as a success rate. The experiment on the (Semantic Matching using instances with WUP >= 0.5) results in 64% as a success rate. The experiment conducted (Semantic Matching using instances with WUP >= 0.6) results in 65% as success rate. The (Semantic Matching using instances with WUP >= 0.7) experiment results in 66% as a success rate. The experiment on the (Semantic Matching using instances with WUP >= 0.8) results in 74% as success rate. Finally, the experiment on the (Semantic Matching using instances with WUP >= 0.9) results in 90 % as success rate.

The experiment from the second method when we using frequency showed (Semantic Matching using concepts only) results in 75% as a success rate. The experiment on the (Semantic Matching using instances with WUP >= 0.5) results in 78% as a success rate. The experiment conducted (Semantic Matching using instances with WUP >= 0.6) results in 79% as success rate. The (Semantic Matching using instances with WUP >= 0.7) experiment results in 79% as a success rate. The experiment on the (Semantic Matching using instances with WUP >= 0.8) results in 83% as success rate. Finally, the experiment on the (Semantic Matching using instances with WUP >= 0.9) results in 90 % as success rate.

The results prove that our technique achieved the highest success rate and enhance the matching results. We attribute this enhancement rate to the usage of our approach which is based on instances and the WUP >= 0.9.

5.5 Results Comparison with related works

Algorithm	Myo, 2003	Pengbin, 2009	Viktor, 2006	Sui, 2012	Bhavana, 2012	Presented
Main Idea	This paper concerns instance detection from a website. It based on link chain extraction and sequential covering	Semantic matching for web pages concepts. The semantic measure is based on semantic distance	Instances extraction from web documents. It is based on redundancy information.	Synchronous attributes based instances extraction. the extraction is contextual similarity based and association between concepts	This paper extracts instances pairs from HTML file, it's based on HTML clustering tables using Hearst patterns	This paper is related to two subjects, the first is to extract instances from vendor and input query, and the second is to find the best match between queries and vendors using semantic matching
Application field	Web page form Yahoo! Movie website	Semantic web services	Art web services	Web services	HTML corpus pages	Renewable energy field
Marking procedure	The user marks the HTML tags manually	No marking	By programmer	By programmer	Automatic marking	The user marks manually one random page, then the marking is updated automatically
Accuracy	73.25%	-----	60%	69%	84.5%	90%

5.6 Main Contributions

There are many contributions in this research which can summarized in the following points:

1- Using semantic matching with instances: in this research we used the instances in the semantic matching. The results showed how can these instances improving the semantic matching.

2- Cutting point for RE domain is 0.9: different threshold cutting point used in this research to determine which is cutting point for the RE domain. The experiments within chapter showed 0.9 is the best threshold for this research.

3- Enhance IDE algorithm by using WUP measure: this research enhanced the IDE algorithm through using WUP similarity measure.

4- We used enhance IDE algorithm in the RE domain: using the IDE algorithm after enhanced in the RE domain to show how this algorithm helps us in the extracting instances process.

5- Extracts items by using KAON ontology building's tool: this research used the KAON ontology building's tool for extracting concepts with different frequency from documents related to RE domain to use it in the semantic matching, and to show the difference between results when we using concepts or instances.

6- Deploy ontology and IDE algorithm: this research deploy the ontology with the enhanced IDE algorithm to automate the extracting instances process and to improve the semantic matching process.

CHAPTER 6

Conclusions and Future work

6.1 Overview

This chapter contains future works and summarizes the work done in this thesis besides conclusion of the thesis. We discuss the future work that will indicate how can exploit this thesis to build new idea and resolve new problems. Conclusion of thesis is based on practical results of experiments conducted in order to check and analyze the system efficiency.

6.2 Conclusion

This thesis concluded RE ontology, KAON ontology building tools for extracting concepts of the RE domain. Additionally, WUP also used in similarity measure in this research. Extracting instances using enhancing IDE algorithm that will help to enhance and improving Semantic matching. After that, to test our algorithm we have semantic similarity measure between extracted instances from RE providers and customer requirements to explain any RE providers meets customer requirements. Also, it can be concluded that attaining customer requirements and needs be done by using our approach, where through this system, it is possible to match the RE customer requirements with RE provider services. The system do this through analyzing RE customer query semantically and calculate its similarity with RE providers using an cumulative process.

6.3 Summary

1- In our research we studied the RE, its ontology concepts, extracting instances and the semantic similarity measure. Also presented how we improving semantic matching using instances.

2- In order to extracting ontology concepts, we collected data (documents, reports, white papers... etc) for RE domain and then converted all documents into text file. Then extract the concepts using KAON ontology building tools.

3- The benefit of the previous step to explain any similarities measuring use in this research.

4- Extracting ontology instances for the 9 providers and 30 query (customer requirements) using modified IDE algorithm.

5- Using these extracted instances to improving semantic matching by retrieving information that are more precise and by satisfies consumers intentions.

6- We have matching between RE providers and customer requirements by semantic similarity measure (by WUP similarity measure) to help customer to decide which RE providers meets their requirements.

6.4 Future Work

Through conducting this research, many ideas and issues were unfolded but not accomplished yet because of time, resources, and other constraints. We would like to suggest a few ideas for future study:

1- Possibility to use our approach to measure other domains semantically.

2- Build full ontology for RE domain.

3- Develop the discovery system for RE providers. This system can work as help administrative to aid users in looking for RE provider. The users in this system search only by using words and the guide system with appropriate RE providers.
4- Make our system online, to let the measure be more accurate and reliable.
5- Using another similarity measure.
6- Extract the level three of instances.
7- Generalize the results to the other domain.

References

Al-bahadili H., Issa G., Al-akhras M., Maqousi A. (2013) The first international conference and exhibition on the application of the information technology in developing renewable energy processes and systems. University of Petra.

Anantharangachar R., Ramani S., Rajagopalan S. (2013) Ontology Guided Information Extraction from Unstructured Text, International Institute of Information Technology. International journal of web and semantic technology (IJWesT) Vol.4, No.1.

Baader F., Sertkaya B., Turhan A. (2005) Computing the Least Common Subsumer w.r.t. a Background Terminology. Theoretical Computer Science, TU Dresden, Germany.

Banerjee S., Pedersen T. (2003) Extended gloss overlaps as a measure of semantic relatedness. In Proceedings of the Eighteenth International Joint Conference on Artificial Intelligence, pages 805–810.

Bellare K., McCallum A. (2007) Learning extractors from unlabeled text using relevant databases. In Proceedings of the Sixth International Workshop on Information Extraction on the Web, pp. 74-84.

Bermejo J. (2007). A Simplified Guide to Create an Ontology. The autonomous systems laboratory, Universidad Politecnica De Madrid. ASlab R-2007-004 v0.1 Draft.

Budanitsky A., Hirst, G. (2006). Evaluating wordnet-based measures of lexical semantic relatedness. Computational Linguistics, 32(1), 13-47.

Cartlidge, E. (2011). Saving for a rainy day. Science (Vol 334). pp. 922–924.

Council on Foreign Relations, (2012). "Public Opinion on Global Issues: Chapter 5b: World Opinion on Energy Security". pp. 13-14.

Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies. Where is their meeting point?. *Data & knowledge engineering*, *46*(1), 41-64.

Dalvi B., William W. Cohen, Jamie Callan, (2011). Extracting Sets of Entities from the Web Using Unsupervised Information Extraction, School of Computer Science, Carnegie Mellon University.

Dai W., You Y., Wang, W., Sun, Y., & Li, T. (2011). Search Engine System Based on Ontology of Technological Resources. *Journal of Software*, *6*(9), 1729-1736.

Danger R., Sanz I., Berlanga, R., Ruíz-Shulcloper, J. (2005) A proposal for the automatic generation of instances from unstructured text University of Oriente, Santiago de Cuba (Cuba) , Universitat Jaume I, Castellón (Spain) ,Institute of Cybernetics, Mathematics and Physics, La Habana (Cuba).

De Boer V., Van Someren M., Wielinga B. J. (2006) Extracting Instances of Relations from Web Documents Using Redundancy, Human-Computer Studies Laboratory, Informatics Institute, Universiteit van Amsterdam.

Etzioni, O., Cafarella, M.J., Downey, D., Popescu, A. M., Shaked, T., Soderland, S., Weld, D. S., and Yates, A.(2005) Unsupervised named-entity extraction from the web: An experimental study. Artif. Intell., 165(1):91{134}.

Fazzinga B., Gianforme G., Gottlob G., and Lukasiewicz T. (2011) Semantic Web search based on ontological conjunctive queries. Web Semantics: Science, Services and Agents on the World Wide Web, 9(4), 453-473.

Fernández-López M. (1999). Overview of methodologies for building ontologies.

Gabel T. Sure Y., Voelker J. (2004). D3. 1.1. a: Kaon-ontology management infrastructure. SEKT informal deliverable. Vol.102, pp. 20-38.

Ganesan V., Swaminathan R., Thenmozhi M.(2012). Similarity Measure Based On Edge Counting Using Ontology. International Journal of Engineering Research and Development.

Giunchiglia F., Yatskevich M., Shvaiko, P. (2008). Semantic Matching: Algorithms and Implementation. Department of Information and Communication Technology University of Trento,38050 Povo, Trento, Italy.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition*, *5*(2), 199-220.

Hearst M. A. (1992) Automatic acquisition of hyponyms from large text corpora. In In Proceedings of the 14th International Conference on Computational Linguistics, pages 539{545}.

Jiang B., Zhu M., Wang J. (2013) Ontology-Based Information Extraction of Crop Diseases on Chinese Web Pages , JOURNAL OF COMPUTERS. Zhejiang Gongshang University, Hangzhou, China.

Jiang, R. (2013). From ontology to semantic similarity: calculation of ontology-based semantic similarity. The Scientific World Journal.

Kayed A., Nizar M., Alfayoumi M. (2010) Ontology Concepts for Requirements Engineering Process in E-Government Applications [Conference]. - [s.l.] : IEEE.

Kayed, A., El-Qawasmeh, E., & Qawaqneh, Z. (2010). Ranking web sites using domain ontology concepts. Information & management, Fahad Bin Sultan University, Saudi Arabia. King Saud University, Saudi Arabia. Jordan University of Science and Technology, Jordan 47(7), 350-355.

Kayed, A., Hirzalla, N., Ahmad, H., & Al Faisal, (2011). Extracting Concepts for Software Components. Trends in Network and Communications, 694-699.

Kozareva Z. (2006) Bootstrapping named entity recognition with automatically generated gazetteer lists. In EACL. The Association for Computer Linguistics.

Kozareva Z., Riloff E., Hovy E. (2008) Semantic class learning from the web with hyponym pattern linkage graphs. In Proceedings of ACL-08: HLT, pages 1048–1056, Columbus, Ohio, June. Association for Computational Linguistics.

Li, Y., Bandar, Z. A., McLean, D. (2003) An approach for measuring semantic similarity between words using multiple information sources. Knowledge and Data Engineering, IEEE Transactions on, 15(4), 871-882.

Lukasiewicz, T., Fazzinga, B., Gianforme, G., Gottlob, G. (2012) Semantic Web Search Based on Ontological Conjunctive Queries. *Web Semantics: Science, Services and Agents on the World Wide Web*, *9*(4).

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.

Mintz M., Bills S., Snow R., Jurafsky D. (2009) Distant supervision for relation extraction without labeled data. In Proceedings of the 47th Annual Meeting of the Association for Computational Linguistics.

Miller, J. (2011). A critical introduction to syntax. Heads. Journal of Linguistics 21, pp. 1-29.

Michelizzi, J. (2005) Semantic relatedness applied to all words sense disambiguation, (Doctoral dissertation), University of Minnesota, Twin Cities, U.S., (On-Line), available: http://www.d.umn.edu/~tpederse/Pubs/jason-thesis.pdf.

Mufson S. (2010) Solar power project in Mojave Desert gets $1.4 billion boost from stimulus funds Washington Post.

Naing M., Lim E., Hoe-Lian Goh D. (2003) On Extracting Link Information of Relationship Instances from a Web Site, Centre for Advanced Information Systems School of Computer Engineering Nanyang Technological University, Nanyang.

Noy N., McGuinness D. (2009) "Ontology Development 101: A Guide to Creating Your First Ontology", [pdf], Stanford, [online]: 4 Jon 2013, available at: http://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html.

Nagwani N., Verma S. (2011) A frequent term and semantic similarity based single document text summarization algorithm, International Journal of Computer Applications. 17 (2), 36-40, Doi: 10.1109/ICTKE.2012.6152388.

Pasca M. (2007). Organizing and searching the world wide web of facts – step two: harnessing the wisdom of the crowds. In WWW '07: Proceedings of the 16th international conference on World Wide Web, pages 101–110, New York, NY, USA. ACM.

Pasca M. (2007) Organizing and searching the world wide web of facts { step two: harnessing the wisdom of the crowds. In WWW '07: Proceedings of the 16thinternational conference on World Wide Web, pages 101{110, New York, NY, USA.

Pearson R. and Eisaman M. (2012) "Energy Storage via Carbon-Neutral Fuels Made From CO2, Water, and Renewable Energy. Proceedings of the IEEE. P. 37-62.

Qiu L., Kan M., Chua T. (2006) Paraphrase recognition via dissimilarity significance classification. In Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, pages 18–26, Sydney, Australia, July. Association for Computational Linguistics.

Richard C. Wang, William W. Cohen, (2008) Automatic Set Instance Extraction using the Web, Carnegie Mellon University.

Richard C. Wang, William W. Cohen (2008) Iterative set expansion of named entities using the web. In ICDM, pages 1091–1096. IEEE Computer Society.

Richard C. Wang, William W. Cohen (2007) Language-independent set expansion of named entities using the web. In ICDM, pages 342–350. IEEE Computer Society.

Rajan, J. M., Lakshmi, M. D. (2012) Ontology-based Semantic Search Engine for Healthcare Services. *International Journal*, *4*.

Seddiqui M. H., Aono M. (2010) Ontology Instance Matching by Considering Semantic Link Cloud, In Proceedings of 9th WSEAS International Conf. on Applications of Computer Engineering.

Shang, S. S., Lin J. S. (2005) A Model for Understanding the Market Orientation Effects of CRM on the Organizational Process.

Sills B. (2011). "Solar May Produce Most of World's Power by 2060, IEA Says. European solar energy association. P. 9.

Snow R., Jurafsky D., Andrew Y. Ng. (2006) Semantic taxonomy induction from heterogenous evidence. In ACL '06: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the ACL, pages 801–808, Morristown, NJ, USA. Association for Computational Linguistics.

Souleiman, H., O'Riain, S., Curry, E. (2012). Approximate Semantic Matching of Heterogeneous Events. In 6th ACM International Conference on Distributed Event-Based Systems (DEBS 2012), 252–263. Berlin, Germany: ACM.

Swartout B., Patil R., Knight K., Russ T. (1996) Toward Distributed Use of Large-Scale Ontologies, In Proc. of the Tenth Workshop on Knowledge Acquisition for Knowledge-Based Systems.

Van Durme B., Pasca M. (2008). Finding Cars, Goddesses and Enzymes: Parametrizable Acquisition of Labeled Instances for Open-Domain Information Extraction, Third AAAI Conference on Artificial Intelligence.

Vadrevu S., Nagarajan S., Gelgi F., Davulcu H. (2004) Automated Metadata and Instance Extraction from News Web Sites, Department of Computer Science and Engineering, Arizona State University,USA.

Wu F., Weld D. (2007). Autonomously semantifying Wikipedia. In: ACM sixteenth conference on Information and Knowledge Management (CIKM), Lisbon, Portugal. Vol. 4273, pp.709-722.

Wang R. C., Schlaefer N., Cohen W., Nyberg E. (2008) Automatic set expansion for list question answering. In Proceedings of the 2008 Conference on Empirical Methods in Natural Language

Processing, pages 947{954, Honolulu, Hawaii, October 2008. Association for Computational Linguistics.

Wu Z., Palmer M. (1994) Verb semantics and lexical selection. In 32nd. Annual Meeting of the Association for Computational Linguistics, pages 133 –138, New Mexico State University, Las Cruces, New Mexico.

Zhifang S., Wei K., Ye T. (2012) Synchronously Extracting Instances and Attributes for the Concepts from the Web. International Journal of Knowledge and Language Processing, Shanghai Research Institute of China Telecom.

Zhang J., Wang Y., Wei H. (2012) An interaction framework of service-oriented ontology learning. In *Proceedings of the 21st ACM international conference on Information and knowledge management*. pp. 2303-2306.

Websites References

1- http://blogs.msdn.com/b/jaredpar/archive/2009/08/26/why-no-linq-in-debugger-windows.aspx.

2- http://ws4jdemo.appspot.com/.

3- http://www.sociologyguide.com/basic-concepts/Social-Distance.php.

4- http://en.wikipedia.org/wiki/Random_measure.

5- http://funsimmat.bioinf.mpi-inf.mpg.de/help3.php.

Appendix:

Table 3-1: Results of extracting ontology concepts for RE

Frequency=5, No. of Concepts=102

Air	Back	Billion	Biodiesel	Biomass	Capacity	Carbon
Cent	Change	Change mitig	Classification	Climate	Climate change	Climate change mitig
Coal	Community	Consumption	Controversion	Conventional gener	Core	Cost
Demand	Develop	Distribution	Earth	Economic	Efficient	Electricity
Electricity gener	End	Energy	Energy storage	Environment	Example	Form
Fossil	Fossil fuel	Fuel	Gas	Gener	Geothermal energy	Government
Grid	Heat	Hydroelectricity	Hydropower	Index	Industrial	Investment
Level	Liquid	Long term	Low carbon	Market	Material	Micro fit
Million	Mitig	Natural gas	Need	Network	Offshore wind	Oil
Operation	Output	Peat	Plant	Plant material	Plant source	Policy
Potential	Power	Product	Program	Project	Quarter	Rang
Rate	Regrowth	Renew	Renewable electricity	Renewable energy	Renewable power	Result
Scale	Second	Share	Solar	Solar energy	Solar power	Source
State	Storage	Sun	Support	Surface	System	Temperature
Term	Time	Transport	Use	Wait	Water	Wind
Wind power	World	year				

Table 3-2: Results of extracting ontology concepts for RE

Frequency=7, No. of Concept=63

Billion	Biomass	Capacity	Carbon	Cent	Change	Classification	Climate
Climate change	Coal	Community	Consumption	Cost	Demand	Develop	Earth
Efficient	Electricity	Electricity gener	End	Energy	Environment	Fossil	Fossil fuel
Fuel	Future	Gas	Gener	Grid	Heat	Industrial	Investment
Market	Million	Oil	Peat	Plant	Power	Product	Program
Project	Quarter	Rate	Renewable energy	Renewable power	Share	Solar	Solar energy
Solar power	Source	State	Sun	Support	System	Temperature	Term
Time	Use	Wait	Water	Wind	World	Year	

Table 3-3: Results of extracting ontology concepts for RE

Frequency=10, No. of Concept=38

Billion	Biomass	Capacity	Cent	Change	Climate	Coal
Develop	Earth	Electricity	Electricity gener	Energy	Fuel	Gener
Grid	Heat	Investment	Million	Oil	Peat	Power
Product	Quarter	Rate	Renewable	Renewable	Share	Source

			energy	power		
System	Use	Water	Wind	World	Year	

Table 3-4: Results of extracting ontology concepts for RE

Frequency=13, No. of Concept=28

Biomass	Capacity	Cent	Change	Climate	Develop	Electricity
Energy	Fuel	Gas	Gener	Heat	Investment	Million
Peat	Plant	Power	Product	Quarter	Rate	Renewable energy
Solar	Source	Use	Water	Wind	World	Year

Table 3-5: Results of extracting ontology concepts for RE

Frequency=15, No. of Concept=21

Biomass	Capacity	Cent	Electricity	Energy	Fuel	Gas	Gener
Heat	Investment	Peat	Plant	Quarter	Renewable energy	Solar	Source
Water	Wind	World	year				

Table 3-6: sample of standard concepts

no.	concept1	concept2	Wup	error	lin	error	path	error
1	car	automobile	100%	0%	100%	0%	100%	0%
2	gem	jewel	100%	0%	100%	0%	100%	0%
3	journey	voyage	95%	5%	82%	18%	50%	50%
4	boy	lad	95%	5%	79%	21%	50%	50%
5	coast	shore	92%	8%	96%	4%	50%	50%
6	asylum	madhouse	95%	5%	98%	2%	50%	50%
7	magician	wizard	100%	0%	100%	0%	100%	0%
8	midday	noon	100%	0%	100%	0%	100%	0%
9	furnace	stove	57%	43%	22%	78%	10%	90%
10	food	fruit	47%	53%	15%	85%	10%	90%
11	bird	cock	95%	5%	78%	22%	50%	50%
12	tool	implement	94%	6%	91%	9%	50%	50%
13	brother	monk	95%	5%	20%	80%	50%	50%
14	crane	implement	77%	23%	33%	67%	20%	80%
15	lad	brother	80%	20%	24%	76%	20%	80%
16	journey	Car	19%	81%	0%	100%	0%	100%
17	monk	oracle	69%	31%	18%	82%	12%	88%
18	food	rooster	28%	72%	0%	100%	0%	100%
19	coast	Hill	71%	29%	72%	28%	20%	80%
20	forest	graveyard	50%	50%	11%	89%	1%	99%
21	monk	slave	80%	20%	20%	80%	20%	80%
22	coast	forest	61%	39%	11%	89%	16%	84%
23	lad	wizard	80%	20%	22%	78%	20%	80%
24	chord	smile	44%	56%	32%	68%	12%	88%
25	glass	magician	53%	47%	14%	86%	12%	88%
26	noon	string	35%	65%	0%	100%	0%	100%

27	rooster	voyage	14%	86%	0%	100%	0%	100%
		Average		29%		54%		66%

1. The extracted instances from providers when WUP >= 0.9

Provider1	Provider2	Provider3	Provider4	Provider5	Provider7	Provider8	Provider9	Provider10
Generation	Wind	Water	Wind	China	Operation	Business	Company	Businesses
Design	Power	Water	Wind	Industry	Generation	Products	Engineering	Countries
Performance	Wind	Design	Offer	Power	Water	construction	Company	Quality
Wind	Production	Installation	Business	Power	Resources	Power	Business	Home
Design	Wind	Assets	Wind	Power	Power	Power	Construction	Technologies
Operations	Generation	Water	Power	Capacity	Businesses	Power	Division	Products
Wind	Wind	Installation	Wind	Water	Power	Company	Comprises	Activity
Production	Generation	Facility	Wind	Businesses	Facilities	Technology	Execution	Company
Wind	Power	Water	Wind	Minutes	Water	Maintenance	Generation	Investments
Capabilities	Transaction	Energy	Wind	Power	Resources	Businesses	Division	Businesses
Wind	Transactions	Water	Power	Power	Engineering	Buildings	Operation	Offer
Wind	Business	Energy	Business	Investment	Resources	Range	Generation	Methods
Wind	Investment	Design	Technology	Home	Capacity	Performance	Power	Combination
Operations	Wind	Water	Business	Performance	Engineering	Installation	Production	Power
Operations	Generation	Water	Business	Products	Design	Power	Comprises	Power
Wind	Facilities	Water	Business	Technologies	Power	Business	Activities	Technology
Design	Wind	Water	Technology	Products	Water	Power	Production	Range
Wind	Wind	Buildings	Business	Design	Division	Power	Technology	Products
Design	Power	Design	Power	Products	Design	Business	Water	Business
Wind	Activities	Installation	Power	Performance	Range	Power	Business	Initiatives
Capacity	Power	Current	Production	Quality	Utilities	Project	Construction	Energy
Generation	Power	Company	Energy	Power	Demonstration	Project	Engineering	Business
Power	Wind	Assets	Company	Products	Grid	Power	Activities	Countries
Power	Energy	Engineering	Wind	China	Capability	Power	Water	Energy
Wind	Wind	Energy	Energy	China	Utilities	Maintenance	Generation	Quality
Power	Energy	Water	Wind	Industry	Capacity	Countries	Water	Energy
Power	Energy	Water	Technology	Industry	Operation	Company	Portfolio	Energy
Wind	Production	Facility	Products	Company	Operation	Power	Assets	Energy
Power	Company	Provide	Offer	Power	Generation	Power	Power	Energy
Wind	Company	Design	Energy	Products	Facilities	Technology	Activity	Household
Current	Wind	Provide	Business	Company	Provision	Businesses	Generation	Energy
Wind	Energy	Installation	Maintenance	Products	Engineering	Buildings	Assets	Activity
Generation	Generation		Wind	Power	Water	Range	Businesses	Company
Capacity	Wind		Power	Energy	Resources	Power	Company	Provision
Technology	Energy		Value	Capacity	Facilities	Energy	Activities	Technology
Design	Generation		Maintenance	Food	Company	Power	Company	Energy
Company	Life		Wind	Water	Energy	Power	Engineering	Energy
Business	Power		Capacity	Energy	Provision	Performance	Company	Energy
Wind	Value		Wind	Businesses	Engineering	Power	Business	Technologies
Engineering	Industry		Wind	Energy	Project	Energy	Engineering	Businesses
Design	Ability		Company	Sake	Company	Installation	Division	Offer
Production	Business		Wind	Energy	Energy	Power	Comprises	Household
Engineering	Investment		Power	Energy	Resources	Power	Generation	Energy
Wind	Lead		Industry	Energy	Energy	Energy	Power	Power
Operation	Investment		Company	Sunlight	Power	Maintenance	Division	Provide
Technologies	Wind		Business	Power	Division	Engineering	Operation	Company
Company	Generation		Technology	Energy	Engineering	Operations	Generation	Power
Operations	Facilities		Business	Businesses	Engineering	Company	Power	Household
Company	Facilities		Business	Power	Businesses	Power	Wind	Energy
Wind	Wind		Business	Grid	Operation	Technology	Production	Technology
Capabilities	Generation		China	Year	Power	Performance	Comprises	Energy
Wind	Facility		Business	Energy	Facilities	Years	Production	Investment
Engineering	Energy		Technology	Years	Provision	Power	Technology	Business
Wind	Wind		Business	Investment	Engineering	Energy	Company	Plan
Operations	Energy		Engineering	Business	Water	Production	Technology	Businesses
Quality	Energy		Engineering	Energy	Resources	Sun	Energy	Energy
Design	Company		Company	Energy	Facilities	Sun	Resources	Range

Wind	Value		China	Home	Company		Water	Home
Installation	Power		Power	Value	Energy		Water	Energy
Performance	Resources		Energy	Energy	Resources		Business	Years
Design	Business		Resources	Quality	Energy		Engineering	Energy
Installation			Energy	Performance	Division		Engineering	Energy
Lift			Energy	Products	Engineering		Engineering	Energy
Generation			Energy	Technologies	Engineering		Engineering	
Grid			Power	Quality	Division		Energy	
Years			Wind	Products	Energy		Water	
			Energy	Project	Division		Years	
			Quality	Engineering	Energy		Generation	
			Life	Design	Resources		Water	
			Energy	Project	Energy		Portfolio	
			Supply	Industry	Facilities		Assets	
			Energy	Products	Company		Power	
			Food	Performance	Capacity		Activity	
			Production	Company	Engineering		Operation	
			Energy	Quality	Division		Wind	
				Field	Engineering		Energy	
				Field	Year		Generation	
					Water		Assets	
					Design		Production	
					Power		Businesses	
					Generation		Technology	
					Facilities		Company	
					Water			
					Engineering			
					Division			
					Design			
					Energy			
					Division			
					Energy			
					Businesses			
					Energy			
					Range			
					Generation			
					Power			
					Operation			
					Facilities			
					Company			
					Energy			
					Energy			
					Energy			
					Supply			
					Capability			
					Company			
					Energy			
					Resources			
					Plan			
					Resources			
					Generation			
					Energy			
					Capacity			
					Operation			
					Grid			
					Grid			
					Power			

2. The extracted instances from queries when WUP >= 0.9

QI 1	QI 2	QI 3	QI 4	QI 5	QI 6	QI 7	QI 8
company	generation	wind	wind	engineering	wind	production	Engineering
power	current	energy	energy	buildings	energy	wind	Wind
energy	energy	solarize	energy	energy	energy	energy	Energy

power	facilities	countries	wave	household	project	power	Field
wind	range	generation	generation	home	china	power	Energy
power	energy	sun	wave	power	power	supply	Power
design	power	power	wind	energy	energy		Value
operations	capabilities	buinesses		resources	company		Business
years		supply		wind	range		
		year		water	value		
		company		sun	cost		

QI 9	QI 10	QI 11	QI 12	QI 13	QI 14	QI 15	QI 16
engineering	company	energy	current	energy	businesses	company	Company
energy	water	water	wind	wind	company	energy	Design
wind	resources	power	energy	generation	wind	power	Wind
company	energy	energy	energy	energy	energy	wind	Generation
might	air	company	company	activities	energy	energy	Resources
offer	current		water	businesses	power	energy	Energy
company	power		power	company	energy	company	Power
	energy		energy		wind	offer	Businesses
	production		buildings		quality	buildings	
	company				company		

QI 17	QI 18	QI 19	QI 20	QI 21	QI 22	QI 23	QI 24
company	wind	energy	china	company	wind	water	company
investment	power	water	power	wind	energy	energy	Wind
wind	energy	energy	water	energy	air	energy	Power
company	products	power	energy	energy	current	facilities	Wind
technologies	provide	technologies	energy	production	energy	water	Energy
products	wind	products	supply	power	power	energy	Energy
	energy	businesses	capacity	company	energy	technologies	Supply
	countries		water	technologies	facilities	power	company
	power		design	place	company	power	Years
	energy			company			
	company						
	Energy						

QI 25	QI 26	QI 27	QI 28	QI 29	QI 30
company	energy	energy	wind	power	company
wind	supply	resources	energy	company	power

water	energy	wind	generation	power	resources
power	power	power	wind	energy	power
energy	wind	production	china	company	energy
power	resources	solarize	company		energy
energy	wind	range			technologies
	design				energy
	wave				range
	company				

3. Semantic Matching between the providers instances and queries instances.

	P1	Human	**Error**	P2	Human	**Error**	P3	Human	**Error**
Q1	86%	70%	**16%**	57%	60%	**3%**	50%	55%	**5%**
Q2	50%	60%	**10%**	63%	70%	**7%**	38%	50%	**12%**
Q3	36%	50%	**14%**	55%	65%	**10%**	18%	40%	**22%**
Q4	50%	55%	**5%**	75%	60%	**15%**	50%	60%	**10%**
Q5	30%	40%	**10%**	40%	55%	**15%**	40%	50%	**10%**
Q6	33%	50%	**17%**	56%	50%	**6%**	78%	55%	**23%**
Q7	50%	45%	**5%**	83%	75%	**8%**	17%	35%	**18%**
Q8	57%	40%	**17%**	71%	60%	**11%**	29%	50%	**21%**
Q9	50%	50%	**0%**	50%	65%	**15%**	50%	65%	**15%**
Q10	50%	65%	**15%**	63%	60%	**3%**	50%	50%	**0%**
Q11	50%	60%	**10%**	75%	75%	**0%**	75%	70%	**5%**
Q12	57%	65%	**8%**	57%	60%	**3%**	71%	70%	**1%**
Q13	50%	60%	**10%**	83%	80%	**3%**	33%	55%	**22%**
Q14	50%	60%	**10%**	67%	60%	**7%**	33%	65%	**32%**
Q15	50%	55%	**5%**	67%	85%	**18%**	50%	60%	**10%**
Q16	71%	60%	**11%**	86%	70%	**16%**	43%	55%	**12%**
Q17	50%	50%	**0%**	50%	65%	**15%**	17%	45%	**28%**
Q18	43%	60%	**17%**	57%	70%	**13%**	43%	65%	**22%**
Q19	33%	50%	**17%**	33%	60%	**27%**	50%	55%	**5%**
Q20	43%	50%	**7%**	29%	55%	**26%**	43%	50%	**7%**
Q21	71%	75%	**4%**	71%	85%	**14%**	14%	40%	**26%**
Q22	57%	60%	**3%**	71%	70%	**1%**	43%	65%	**22%**
Q23	40%	50%	**10%**	60%	75%	**15%**	40%	65%	**25%**
Q24	67%	60%	**7%**	67%	65%	**2%**	50%	60%	**10%**
Q25	60%	65%	**5%**	80%	75%	**5%**	60%	65%	**5%**
Q26	50%	60%	**10%**	50%	65%	**15%**	37%	55%	**18%**
Q27	43%	55%	**12%**	71%	60%	**11%**	14%	35%	**21%**
Q28	60%	60%	**0%**	80%	75%	**5%**	40%	50%	**10%**
Q29	33%	70%	**37%**	100%	85%	**15%**	67%	75%	**8%**
Q30	50%	50%	**0%**	67%	70%	**3%**	33%	60%	**27%**

P4	Human	**Error**	P5	Huamn	**Error**	P6	Human	**Error**
50%	65%	**15%**	57%	55%	**2%**	50%	50%	**0%**
38%	55%	**17%**	38%	50%	**12%**	88%	75%	**13%**
55%	40%	**15%**	45%	55%	**10%**	55%	60%	**5%**
50%	50%	**0%**	50%	55%	**5%**	75%	50%	**25%**
50%	65%	**15%**	50%	65%	**15%**	50%	45%	**5%**
67%	65%	**2%**	67%	65%	**2%**	56%	60%	**4%**
83%	75%	**8%**	33%	55%	**22%**	67%	65%	**2%**
86%	70%	**16%**	86%	75%	**11%**	43%	40%	**3%**
83%	70%	**13%**	50%	65%	**15%**	50%	60%	**10%**
63%	55%	**8%**	50%	70%	**20%**	63%	75%	**12%**
75%	65%	**10%**	100%	80%	**20%**	100%	85%	**15%**
57%	55%	**2%**	57%	75%	**18%**	57%	60%	**3%**
50%	65%	**15%**	50%	70%	**20%**	67%	65%	**2%**
50%	60%	**10%**	83%	75%	**8%**	67%	65%	**2%**
83%	70%	**13%**	50%	75%	**25%**	50%	60%	**10%**
71%	60%	**11%**	57%	70%	**13%**	86%	75%	**11%**
33%	60%	**27%**	83%	80%	**3%**	33%	50%	**17%**
71%	70%	**1%**	57%	75%	**18%**	43%	60%	**17%**
50%	50%	**0%**	100%	85%	**15%**	67%	65%	**2%**
71%	60%	**11%**	86%	80%	**6%**	86%	80%	**6%**
71%	60%	**11%**	57%	65%	**8%**	43%	55%	**12%**
57%	55%	**2%**	43%	65%	**22%**	57%	65%	**8%**
60%	65%	**5%**	20%	75%	**55%**	40%	75%	**35%**
83%	75%	**8%**	67%	60%	**7%**	67%	60%	**7%**
80%	70%	**10%**	80%	75%	**5%**	80%	80%	**0%**
75%	65%	**10%**	50%	60%	**10%**	63%	55%	**8%**
57%	65%	**8%**	29%	50%	**21%**	57%	50%	**7%**
80%	70%	**10%**	60%	65%	**5%**	60%	40%	**20%**
100%	80%	**20%**	100%	85%	**15%**	100%	80%	**20%**
67%	65%	**2%**	67%	85%	**18%**	83%	85%	**2%**

P7	Human	**Error**	P8	Human	**Error**	P9	Human	**Error**
71%	65%	**6%**	71%	70%	**1%**	57%	60%	**3%**
50%	40%	**10%**	50%	50%	**0%**	50%	50%	**0%**
55%	60%	**5%**	55%	60%	**5%**	45%	50%	**5%**
25%	20%	**5%**	100%	90%	**10%**	25%	40%	**15%**
50%	50%	**0%**	60%	65%	**5%**	40%	55%	**15%**
56%	65%	**9%**	44%	55%	**11%**	44%	60%	**16%**
50%	55%	**5%**	67%	70%	**3%**	33%	45%	**12%**

57%	60%	**3%**	71%	60%	**11%**	43%	60%	**17%**
50%	55%	**5%**	67%	65%	**2%**	50%	55%	**5%**
50%	50%	**0%**	63%	60%	**3%**	38%	50%	**12%**
75%	75%	**0%**	100%	90%	**10%**	75%	75%	**0%**
57%	60%	**3%**	71%	70%	**1%**	43%	60%	**17%**
50%	50%	**0%**	100%	90%	**10%**	50%	55%	**5%**
67%	60%	**7%**	50%	75%	**25%**	83%	85%	**2%**
67%	70%	**3%**	50%	70%	**20%**	67%	70%	**3%**
43%	40%	**3%**	86%	80%	**6%**	43%	50%	**7%**
50%	55%	**5%**	50%	60%	**10%**	83%	75%	**8%**
71%	70%	**1%**	57%	80%	**23%**	85%	70%	**15%**
67%	60%	**7%**	67%	65%	**2%**	83%	80%	**3%**
29%	40%	**11%**	29%	50%	**21%**	29%	40%	**11%**
57%	75%	**18%**	71%	70%	**1%**	57%	60%	**3%**
43%	60%	**17%**	43%	60%	**17%**	43%	45%	**2%**
40%	50%	**10%**	20%	50%	**30%**	40%	55%	**15%**
67%	65%	**2%**	83%	80%	**3%**	67%	60%	**7%**
60%	75%	**15%**	100%	85%	**15%**	60%	65%	**5%**
38%	50%	**12%**	63%	65%	**2%**	38%	30%	**8%**
57%	60%	**3%**	71%	75%	**4%**	43%	35%	**8%**
40%	30%	**10%**	80%	75%	**5%**	40%	30%	**10%**
100%	85%	**15%**	100%	90%	**10%**	100%	80%	**20%**
33%	60%	**27%**	67%	80%	**13%**	50%	75%	**25%**

4. The extracted instances from providers when WUP >= 0.8

Provider1	**Provider2**	**Provider3**	**Provider4**	**Provider5**
Generation	Wind	Water	Wind	China
Design	Power	Water	Wind	Industry
Performance	Wind	Design	Offer	Power
Wind	Production	Installation	Business	Power
Design	Wind	Assets	Wind	Power
Operations	Generation	Water	Power	Capacity
Wind	Wind	Installation	Wind	Water
Production	Generation	Facility	Wind	Businesses
Wind	Power	Water	Wind	Minutes
Capabilities	Transaction	Energy	Wind	Power
Wind	Transactions	Water	Power	Power
Wind	Business	Energy	Business	Investment
Wind	Investment	Design	Technology	Home
Operations	Wind	Water	Business	Performance
Operations	Generation	Water	Business	Products

Wind	Facilities	Water	Business	Technologies
Design	Wind	Water	Technology	Products
Wind	Wind	Buildings	Business	Design
Design	Power	Design	Power	Products
Wind	Activities	Installation	Power	Performance
Capacity	Power	Current	Production	Quality
Generation	Power	Company	Energy	Power
Power	Wind	Assets	Company	Products
Power	Energy	Engineering	Wind	China
Wind	Wind	Energy	Energy	China
Power	Energy	Water	Wind	Industry
Power	Energy	Water	Technology	Industry
Wind	Production	Facility	Products	Company
Power	Company	Provide	Offer	Power
Wind	Company	Design	Energy	Products
Current	Wind	Provide	Business	Company
Wind	Energy	installation	Maintenance	Products
Generation	Generation	manufacture	Wind	Power
Capacity	Wind	source	Power	Energy
Technology	Energy	systems	Value	Capacity
Design	Generation	december	Maintenance	Food
Company	Life	source	Wind	Water
Business	Power	systems	Capacity	Energy
Wind	Value	distribution	Wind	Businesses
Engineering	Industry	pacific	Wind	Energy
Design	Ability	heating	Company	Sake
Production	Business	cooling	Wind	Energy
Engineering	Investment	systems	Power	Energy
Wind	Lead	coductivity	Industry	Energy
Operation	Investment	assistance	Company	Sunlight
Technologies	Wind	limit	Business	Power
Company	Generation	draw	Technology	Energy
Operations	Facilities	start	Business	Businesses
Company	Facilities	january	Business	Power
Wind	Wind	footprint	Business	Grid
Capabilities	Generation	time	China	Year
Wind	Facility	use	Business	Energy
Engineering	Energy	free	Technology	Years
Wind	Wind	heating	Business	Investment
Operations	Energy	cooling	Engineering	Business
Quality	Energy	yard	Engineering	Energy
Design	Company	source	Company	Energy
Wind	Value	heat	China	Home
Installation	Power	pump	Power	Value

Performance	Resources	system	Energy	Energy
Design	Business	state	Resources	Quality
Installation	play	art	Energy	Performance
Lift	asset	equipment	Energy	Products
Generation	Life	people	Energy	Technologies
Grid	provider	check	Power	Quality
Years	term	start	Wind	Products
electricity	cash	earth	Energy	Project
close	appreciate		Quality	Engineering
increase	time		Life	Design
assembly	strategy		Energy	Project
manufacturing	combine		Supply	Industry
development	Industry		Energy	Products
maintenance	expertise		Food	Performance
include	Ability		Production	Company
control	capital		Energy	Quality
system	order		superior	Field
development	secure		cost	Field
manufacture	benefit		services	limited
sale	development		works	limited
north	strategy		close	manufacturer
south	platform		partnership	distribution
america	near		core	network
manufacturing	march		development	europe
development	interest		manufacturing	north
maintenance	western		sale	america
management	may		cover	asia
weight	strategy		chain	covering
weight	high		site	stock
light	capital		service	exchange
manufacture	allocation		market	manufacture
transport	build		leading	system
maintenance	buy		position	integration
transport	understanding		day	leader
connection	public		demand	services
standard	safety		systems	standard
size	values		manufacturing	output
connection	partnership		service	markets
	open		area	number
	communication		support	germany
	manage		asia	spain
	ways		europe	italy
	ensure		mediterranean	system
	development		europe	december

	appraoch		manufacturing	manufacturing
			area	rapid
			manufacturing	development
			assembly	use
			service	world
			area	economy
			platform	dioxide
			management	climate
			services	change
			december	balance
			network	even
			germany	fuel
			italy	rise
			spain	sake
			sweden	environment
			norway	finding
			use	today
			pressure	world
			change	source
			way	world
			produce	reach
			produce	germany
			source	italy
			catalyst	reach
			better	cleaner
			stable	family
			essential	electricity
			part	costs
			socirty	increase
			major	top
			challenge	high
			world	world
			today	setting
			security	third
			climate	high
			change	test
			food	time
			access	financing
			strengthening	track
			equity	step
				module
				deployment

Provider7	**Provider8**	**Provider9**	**Provider10**
Operation	Business	Company	Businesses
Generation	Products	Engineering	Countries
Water	construction	Company	Quality
Resources	Power	Business	Home
Power	Power	Construction	Technologies
businesses	Power	Division	Products
Power	Company	Comprises	Activity
Facilities	Technology	Execution	Company
Water	Maintenance	Generation	Investments
Resources	businesses	Division	Businesses
Engineering	Buildings	Operation	Offer
Resources	Range	Generation	Methods
Capacity	Performance	Power	Combination
Engineering	Installation	Production	Power
Design	Power	Comprises	Power
Power	Business	Activities	Technology
Water	Power	Production	Range
Division	Power	Technology	Products
Design	Business	Water	Business
Range	Power	Business	Initiatives
Utilities	Project	Construction	Energy
Demonstration	Project	Engineering	Business
Grid	Power	Activities	Countries
Capability	Power	Water	Energy
Utilities	Maintenance	Generation	Quality
Capacity	Countries	Water	Energy
Operation	Company	Portfolio	Energy
Operation	Power	Assets	Energy
Generation	Power	Power	Energy
Facilities	Technology	Activity	Household
Provision	Businesses	Generation	Energy
Engineering	Buildings	Assets	Activity
Water	Range	businesses	Company
Resources	Power	Company	Provision
Facilities	Energy	Activities	Technology
Company	Power	Company	Energy
Energy	Power	Engineering	Energy
Provision	Performance	Company	Energy
Engineering	Power	Business	Technologies
Project	Energy	Engineering	Businesses
Company	Installation	Division	Offer
Energy	Power	Comprises	Household
Resources	Power	Generation	Energy

Energy	Energy	Power	Power
Power	Maintenance	Division	Provide
Division	Engineering	Operation	Company
Engineering	Operations	Generation	Power
Engineering	Company	Power	Household
Businesses	Power	Wind	Energy
Operation	Technology	Production	Technology
Power	Performance	Comprises	Energy
Facilities	Years	Production	Investment
Provision	Power	Technology	Business
Engineering	Energy	Company	Plan
Water	Production	Technology	Businesses
Resources	Sun	Energy	Energy
Facilities	Sun	Resources	Range
Company	systems	Water	Home
Energy	development	Water	Energy
Resources	plant	Business	Years
Energy	plant	Engineering	Energy
Division	january	Engineering	Energy
Engineering	november	Engineering	Energy
Engineering	history	Engineering	Limited
Division	craftsmanship	Energy	Works
Energy	fact	Water	Increase
Division	cell	Years	People
Energy	percent	Generation	Date
Resources	conversion	Water	Million
Energy	cell	Portfolio	People
Facilities	system	Assets	Million
Company	data	Power	Million
Capacity	utility	Activity	Support
Engineering	services	Operation	Development
Division	equipment	Wind	Limited
Engineering	small	Energy	Hardware
Year	market	Generation	Equipment
Water	dealer	Assets	Performing
Design	network	Production	Necessary
Power	direct	Businesses	Light
Generation	europe	Technology	System
Facilities	systems	Company	Development
Water	solar	spain	Work
Engineering	public	sector	Meet
Division	solar	main	Help
Design	projects	station	Advice
Energy	long	transmission	Impetus

Division	term	concession	Date
Energy	electricity	waste	Million
Businesses	solar	treatment	People
Energy	better	development	Result
Range	solar	may	Million
Generation	real	environment	Use
Power	world	electricity	Million
Operation	world	sea	Support
Facilities	record	market	Development
Company	today	experience	World
Energy	high	gas	Africa
Energy	world	large	
Energy	record	scale	
Supply	level	transmission	
Capability	average	concession	
Company	percent	concession	
Energy	real	long	
Resources	world	term	
Plan	long	pay	
Resources	term	purchase	
Generation	high	transmission	
Energy	percent	demand	
Capacity	plus	focus	
Operation	low	high	
Grid	uptime	development	
Grid		leadership	
Power		position	
development		practice	
ownership			
distribution			
services			
services			
include			
funding			
development			
management			
services			
cover			
development			
ownership			
public			
stock			
exchange			
subsidiary			

owner			
operator			
columbia			
history			
services			
control			
use			
planning			
meet			
utility			
high			
remote			
program			
assist			
remote			
financing			
funding			
access			
capital			
markets			
term			
formation			
capital			
major			
utility			
selection			
near			
columbia			
connection			
north			
michigan			
line			
end			
interest			

5. The extracted instances from queries when WUP >= 0.8.

QI 1	QI 2	QI 3	QI 4	QI 5	QI 6
company	generation	wind	wind	engineering	wind
power	current	energy	energy	buildings	energy
energy	energy	solarize	energy	energy	energy
power	facilities	countries	wave	household	project
wind	range	generation	generation	home	china

power	energy	sun	wave	power	power
design	power	power	wind	energy	energy
operations	capabilities	buinesses	type	resources	company
years	high	supply	fuel	wind	range
require	high	year	speed	water	value
high	increase	company	type	sun	cost
amount	area	require		high	type
area	second			specification	average
type	limit			nature	consumption
matter	limit				area
average	float				
type					

QI 7	QI 8	QI 9	QI 10	QI 11	QI 12
production	engineering	engineering	company	energy	current
wind	wind	energy	water	water	wind
energy	energy	wind	resources	power	energy
power	field	company	energy	energy	energy
power	energy	might	air	company	company
supply	power	offer	current	require	water
require	value	company	power	type	power
average	business	require	energy	average	energy
amount	land	type	production	consumption	buildings
process		plants	company	float	type
type		services	average	area	average
		major	type	support	consumption
			area	type	type
			require		
			support		
			type		

QI 13	QI 14	QI 15	QI 16	QI 17	QI 18
energy	businesses	company	company	company	wind
wind	company	energy	design	investment	power
generation	wind	power	wind	wind	energy
energy	energy	wind	generation	company	products
activities	energy	energy	resources	technologies	provide
businesses	power	energy	energy	products	wind
company	energy	company	power	services	energy
type	wind	offer	businesses	support	countries
area	quality	buildings	average	cover	power

support	company	type	consumption	area	energy
type	type	average		type	company
	type	consumption		propose	energy
	close	require			manufacturer
	area	area			type
	type	distance			average
		type			type

QI 19	QI 20	QI 21	QI 22	QI 23	QI 24
energy	china	company	wind	water	company
water	power	wind	energy	energy	wind
energy	water	energy	air	energy	power
power	energy	energy	current	facilities	wind
technologies	energy	production	energy	water	energy
products	supply	power	power	energy	energy
businesses	capacity	company	energy	technologies	supply
type	water	technologies	facilities	power	company
average	design	place	company	power	years
limit	region	company	support	type	type
set	type	type	type	set	area
type	plus	amount	consumption	electricity	support
	area	support	type	use	type
	support	equipment		include	
	type	near		plants	
				plants	
				plants	
				average	
				consumption	
				type	

QI 25	QI 26	QI 27	QI 28	QI 29	QI 30
company	energy	energy	wind	power	company
wind	supply	resources	energy	company	power
water	energy	wind	generation	power	resources
power	power	power	wind	energy	power
energy	wind	production	china	company	energy
power	resources	solarize	company	station	energy
energy	wind	range	type	infrastructure	technologies
group	design	heat	require	transmission	energy

type	wave	electricity	fuel	average	range
estimate	company	footprint	vehicle	consumption	development
type	type	environment	type	area	alternative
area	area	electricity		support	increase
support	support			type	type
	type				

6. Human Experts

	P1	P2	P3	P4	P5	P6	P7	P8	P9
Q1	70%	60%	55%	65%	55%	50%	65%	70%	60%
Q2	60%	70%	50%	55%	50%	75%	40%	50%	50%
Q3	50%	65%	40%	40%	55%	60%	60%	60%	50%
Q4	55%	60%	60%	50%	55%	50%	20%	90%	40%
Q5	40%	55%	50%	65%	65%	45%	50%	65%	55%
Q6	50%	50%	55%	65%	65%	60%	65%	55%	60%
Q7	45%	75%	35%	75%	55%	65%	55%	70%	45%
Q8	40%	60%	50%	70%	75%	40%	60%	60%	60%
Q9	50%	65%	65%	70%	65%	60%	55%	65%	55%
Q10	65%	60%	50%	55%	70%	75%	50%	60%	50%
Q11	60%	75%	70%	65%	80%	85%	75%	90%	75%
Q12	65%	60%	70%	55%	75%	60%	60%	70%	60%
Q13	60%	80%	55%	65%	70%	65%	50%	90%	55%
Q14	60%	60%	65%	60%	75%	65%	60%	75%	85%
Q15	55%	85%	60%	70%	75%	60%	70%	70%	70%
Q16	60%	70%	55%	60%	70%	75%	40%	80%	50%
Q17	50%	65%	45%	60%	80%	50%	55%	60%	75%
Q18	60%	70%	65%	70%	75%	60%	70%	80%	70%
Q19	50%	60%	55%	50%	85%	65%	60%	65%	80%
Q20	50%	55%	50%	60%	80%	80%	40%	50%	40%
Q21	75%	85%	40%	60%	65%	55%	75%	70%	60%
Q22	60%	70%	65%	55%	65%	65%	60%	60%	45%
Q23	50%	75%	65%	65%	75%	75%	50%	50%	55%
Q24	60%	65%	60%	75%	60%	60%	65%	80%	60%
Q25	65%	75%	65%	70%	75%	80%	75%	85%	65%
Q26	60%	65%	55%	65%	60%	55%	50%	65%	30%
Q27	55%	60%	35%	65%	50%	50%	60%	75%	35%
Q28	60%	75%	50%	70%	65%	40%	30%	75%	30%
Q29	70%	85%	75%	80%	85%	80%	85%	90%	80%
Q30	50%	70%	60%	65%	85%	85%	60%	80%	75%

7. Semantic Matching between above providers instances and queries instances.

	P1	Human	**Error**	P2	Human	**Error**	P3	Human	**Error**	P4	Human	**Error**
Q 1	47%	70%	**23%**	41%	60%	**19%**	18%	55%	**37%**	35%	65%	**30%**
Q 2	31%	60%	**29%**	44%	70%	**26%**	31%	50%	**19%**	25%	55%	**30%**
Q 3	33%	50%	**17%**	50%	65%	**15%**	17%	40%	**23%**	50%	40%	**10%**
Q 4	27%	55%	**28%**	45%	60%	**15%**	27%	60%	**33%**	36%	50%	**14%**
Q 5	21%	40%	**19%**	43%	55%	**12%**	36%	50%	**14%**	43%	65%	**22%**
Q 6	20%	50%	**30%**	47%	50%	**3%**	27%	55%	**28%**	60%	65%	**5%**
Q 7	33%	45%	**12%**	50%	75%	**25%**	8%	35%	**27%**	50%	75%	**25%**
Q 8	44%	40%	**4%**	67%	60%	**7%**	33%	50%	**17%**	78%	70%	**8%**
Q 9	33%	50%	**17%**	33%	65%	**32%**	33%	65%	**32%**	67%	70%	**3%**
Q 10	31%	65%	**34%**	44%	60%	**16%**	38%	50%	**12%**	50%	55%	**5%**
Q 11	15%	60%	**45%**	31%	75%	**44%**	31%	70%	**39%**	46%	65%	**19%**
Q 12	31%	65%	**34%**	46%	60%	**14%**	54%	70%	**16%**	46%	55%	**9%**
Q 13	27%	60%	**33%**	55%	80%	**25%**	27%	55%	**28%**	45%	65%	**20%**
Q 14	40%	60%	**20%**	47%	60%	**13%**	33%	65%	**32%**	60%	60%	**0%**
Q 15	25%	55%	**30%**	44%	85%	**41%**	38%	60%	**22%**	56%	70%	**14%**
Q 16	56%	60%	**4%**	67%	70%	**3%**	33%	55%	**22%**	56%	60%	**4%**
Q 17	30%	50%	**20%**	30%	65%	**35%**	15%	45%	**30%**	54%	60%	**6%**
Q 18	31%	60%	**29%**	56%	70%	**14%**	38%	65%	**27%**	63%	70%	**7%**
Q 19	17%	50%	**33%**	25%	60%	**35%**	33%	55%	**22%**	33%	50%	**17%**
Q 20	20%	50%	**30%**	20%	55%	**35%**	33%	50%	**17%**	53%	60%	**7%**
Q 21	47%	75%	**28%**	60%	85%	**25%**	20%	40%	**20%**	60%	60%	**0%**
Q 22	30%	60%	**30%**	54%	70%	**16%**	38%	65%	**27%**	54%	55%	**1%**
Q 23	25%	50%	**25%**	30%	75%	**45%**	25%	65%	**40%**	30%	65%	**35%**
Q 24	46%	60%	**14%**	54%	65%	**11%**	30%	60%	**30%**	77%	75%	**2%**
Q 25	30%	65%	**35%**	46%	75%	**29%**	30%	65%	**35%**	62%	70%	**8%**
Q 26	36%	60%	**24%**	43%	65%	**22%**	29%	55%	**26%**	71%	65%	**6%**
Q 27	42%	55%	**13%**	42%	60%	**18%**	25%	35%	**10%**	33%	65%	**32%**
Q 28	36%	60%	**24%**	45%	75%	**30%**	18%	50%	**32%**	45%	70%	**25%**
Q 29	30%	70%	**40%**	38%	85%	**47%**	23%	75%	**52%**	54%	80%	**26%**
Q 30	46%	50%	**4%**	62%	70%	**8%**	30%	60%	**30%**	62%	65%	**3%**

P5	Human	**Error**	P6	Human	**Error**	P7	Human	**Error**	P8	Human	**Error**	p9	Human	**Error**
41%	55%	**14%**	35%	50%	**15%**	53%	65%	**12%**	47%	70%	**23%**	35%	60%	**25%**
38%	50%	**12%**	56%	75%	**19%**	38%	40%	**2%**	38%	50%	**12%**	31%	50%	**19%**

42%	55%	**13%**	50%	60%	**10%**	50%	60%	**10%**	50%	60%	**10%**	42%	50%	**8%**
36%	55%	**19%**	36%	50%	**14%**	18%	20%	**2%**	55%	90%	**35%**	18%	40%	**22%**
50%	65%	**15%**	50%	45%	**5%**	50%	50%	**0%**	57%	65%	**8%**	36%	55%	**19%**
53%	65%	**12%**	47%	60%	**13%**	53%	65%	**12%**	40%	55%	**15%**	40%	60%	**20%**
25%	55%	**30%**	42%	65%	**23%**	42%	55%	**13%**	42%	70%	**28%**	25%	45%	**20%**
78%	75%	**3%**	44%	40%	**4%**	56%	60%	**4%**	67%	60%	**7%**	44%	60%	**16%**
42%	65%	**23%**	50%	60%	**10%**	42%	55%	**13%**	42%	65%	**23%**	33%	55%	**22%**
38%	70%	**32%**	44%	75%	**31%**	44%	50%	**6%**	44%	60%	**16%**	38%	50%	**12%**
38%	80%	**42%**	38%	85%	**47%**	38%	75%	**37%**	38%	90%	**52%**	38%	75%	**37%**
46%	75%	**29%**	46%	60%	**14%**	54%	60%	**6%**	54%	70%	**16%**	38%	60%	**22%**
36%	70%	**34%**	45%	65%	**20%**	36%	50%	**14%**	64%	90%	**26%**	45%	55%	**10%**
53%	75%	**22%**	47%	65%	**18%**	47%	60%	**13%**	47%	75%	**28%**	53%	85%	**32%**
38%	75%	**37%**	38%	60%	**22%**	50%	70%	**20%**	38%	70%	**32%**	44%	70%	**26%**
44%	70%	**26%**	78%	75%	**3%**	33%	40%	**7%**	67%	80%	**13%**	33%	50%	**17%**
62%	80%	**18%**	38%	50%	**12%**	38%	55%	**17%**	30%	60%	**30%**	54%	75%	**21%**
50%	75%	**25%**	44%	60%	**16%**	63%	70%	**7%**	56%	80%	**24%**	63%	70%	**7%**
58%	85%	**27%**	42%	65%	**23%**	50%	60%	**10%**	42%	65%	**23%**	50%	80%	**30%**
53%	80%	**27%**	53%	80%	**27%**	20%	40%	**20%**	27%	50%	**23%**	27%	40%	**13%**
47%	65%	**18%**	47%	55%	**8%**	53%	75%	**22%**	53%	70%	**17%**	53%	60%	**7%**
38%	65%	**27%**	46%	65%	**19%**	38%	60%	**22%**	38%	60%	**22%**	46%	45%	**1%**
50%	75%	**25%**	50%	75%	**25%**	35%	50%	**15%**	25%	50%	**25%**	35%	55%	**20%**
46%	60%	**14%**	46%	60%	**14%**	46%	65%	**19%**	62%	80%	**18%**	54%	60%	**6%**
46%	75%	**29%**	46%	80%	**34%**	38%	75%	**37%**	54%	85%	**31%**	46%	65%	**19%**
36%	60%	**24%**	43%	55%	**12%**	29%	50%	**21%**	50%	65%	**15%**	36%	30%	**6%**
42%	50%	**8%**	33%	50%	**17%**	50%	60%	**10%**	67%	75%	**8%**	25%	35%	**10%**
36%	65%	**29%**	27%	40%	**13%**	18%	30%	**12%**	45%	75%	**30%**	18%	30%	**12%**
38%	85%	**47%**	38%	80%	**42%**	46%	85%	**39%**	54%	90%	**36%**	46%	80%	**34%**
69%	85%	**16%**	69%	85%	**16%**	46%	60%	**14%**	62%	80%	**18%**	77%	75%	**2%**

8. The extracted instances from providers when WUP >= 0.7

Provider1	**Provider2**	**Provider3**	**Provider4**	**Provider5**
Generation	Wind	Water	Wind	China
Design	Power	Water	Wind	Industry
Performance	Wind	Design	Offer	Power
Wind	Production	Installation	Business	Power
Design	Wind	Assets	Wind	Power
Operations	Generation	Water	Power	Capacity
Wind	Wind	Installation	Wind	Water
Production	Generation	Facility	Wind	Businesses
Wind	Power	Water	Wind	Minutes
Capabilities	Transaction	Energy	Wind	Power

Wind	Transactions	Water	Power	Power
Wind	Business	Energy	Business	Investment
Wind	Investment	Design	Technology	Home
Operations	Wind	Water	Business	Performance
Operations	Generation	Water	Business	Products
Wind	Facilities	Water	Business	Technologies
Design	Wind	Water	Technology	Products
Wind	Wind	Buildings	Business	Design
Design	Power	Design	Power	Products
Wind	Activities	Installation	Power	Performance
Capacity	Power	Current	Production	Quality
Generation	Power	Company	Energy	Power
Power	Wind	Assets	Company	Products
Power	Energy	Engineering	Wind	China
Wind	Wind	Energy	Energy	China
Power	Energy	Water	Wind	Industry
Power	Energy	Water	Technology	Industry
Wind	Production	Facility	Products	Company
Power	Company	Provide	Offer	Power
Wind	Company	Design	Energy	Products
Current	Wind	Provide	Business	Company
Wind	Energy	installation	Maintenance	Products
Generation	Generation	manufacture	Wind	Power
Capacity	Wind	source	Power	Energy
Technology	Energy	systems	Value	Capacity
Design	Generation	december	Maintenance	Food
Company	Life	source	Wind	Water
Business	Power	systems	Capacity	Energy
Wind	Value	distribution	Wind	Businesses
Engineering	Industry	pacific	Wind	Energy
Design	Ability	heating	Company	Sake
Production	Business	cooling	Wind	Energy
Engineering	Investment	systems	Power	Energy
Wind	Lead	coductivity	Industry	Energy
Operation	Investment	assistance	Company	Sunlight
Technologies	Wind	limit	Business	Power
Company	Generation	draw	Technology	Energy
Operations	Facilities	start	Business	Businesses
Company	Facilities	january	Business	Power
Wind	Wind	footprint	Business	Grid
Capabilities	Generation	time	China	Year
Wind	Facility	use	Business	Energy
Engineering	Energy	free	Technology	Years
Wind	Wind	heating	Business	Investment

Operations	Energy	cooling	Engineering	Business
Quality	Energy	yard	Engineering	Energy
Design	Company	source	Company	Energy
Wind	Value	heat	China	Home
Installation	Power	pump	Power	Value
Performance	Resources	system	Energy	Energy
Design	Business	state	Resources	Quality
Installation	play	art	Energy	Performance
Lift	asset	equipment	Energy	Products
Generation	Life	people	Energy	Technologies
Grid	provider	check	Power	Quality
Years	term	start	Wind	Products
electricity	cash	earth	Energy	Project
close	appreciate	testing	Quality	Engineering
increase	time	devices	Life	Design
assembly	strategy	inventory	Energy	Project
manufacturing	combine	found	Supply	Industry
development	Industry	back	Energy	Products
maintenance	expertise	save	Food	Performance
include	Ability	costs	Production	Company
control	capital	check	Energy	Quality
system	order		superior	Field
development	secure		cost	Field
manufacture	benefit		services	limited
sale	development		works	limited
north	strategy		close	manufacturer
south	platform		partnership	distribution
america	near		core	network
manufacturing	march		development	europe
development	interest		manufacturing	north
maintenance	western		sale	america
management	may		cover	asia
weight	strategy		chain	covering
weight	high		site	stock
light	capital		service	exchange
manufacture	allocation		market	manufacture
transport	build		leading	system
maintenance	buy		position	integration
transport	understanding		day	leader
connection	public		demand	services
standard	safety		systems	standard
size	values		manufacturing	output

connection	partnership		service	markets
five	open		area	number
farm	communication		support	germany
segment	manage		asia	spain
segment	ways		europe	italy
africa	ensure		mediterranean	system
main	development		europe	december
variety	appraoch		manufacturing	manufacturing
farm	play		area	rapid
low	primary		manufacturing	development
play	vehicle		assembly	use
primary	one		service	world
vehicle	arizona		area	economy
	california		platform	dioxide
	stable		management	climate
	over		services	change
	focus		december	balance
	california		network	even
	subsidary		germany	fuel
	focus		italy	rise
	accept		spain	sake
	natural		sweden	environment
			norway	finding
			use	today
			pressure	world
			change	source
			way	world
			produce	reach
			produce	germany
			source	italy
			catalyst	reach
			better	cleaner
			stable	family
			essential	electricity
			part	costs
			socirty	increase
			major	top
			challenge	high
			world	world
			today	setting
			security	third
			climate	high

			change	test
			food	time
			access	financing
			strengthening	track
			equity	step
			independence	module
			total	deployment
			bringing	silicon
			par	pioneer
			gas	shaper
			growing	european
			denmark	plant
			active	phoenix
			four	CO2
			six	natural
			central	crisis
			denmark	worsening
			sustainability	pollution
			growing	crisis
			planet	modern
			great	inequality
			founding	facing
			infrastructure	ready
			opportunity	safe
			sustainability	like
			strengthening	california
			poverty	possible
				reduce
				bill
				indendence
				rising
				record
				party
				testing
				stand
				track
				silicon
				step

Provider7	**Provider8**	**Provider9**	**Provider10**
Operation	Business	Company	Businesses
Generation	Products	Engineering	Countries
Water	construction	Company	Quality

Resources	Power	Business	Home
Power	Power	Construction	Technologies
businesses	Power	Division	Products
Power	Company	Comprises	Activity
Facilities	Technology	Execution	Company
Water	Maintenance	Generation	Investments
Resources	businesses	Division	Businesses
Engineering	Buildings	Operation	Offer
Resources	Range	Generation	Methods
Capacity	Performance	Power	Combination
Engineering	Installation	Production	Power
Design	Power	Comprises	Power
Power	Business	Activities	Technology
Water	Power	Production	Range
Division	Power	Technology	Products
Design	Business	Water	Business
Range	Power	Business	Initiatives
Utilities	Project	Construction	Energy
Demonstration	Project	Engineering	Business
Grid	Power	Activities	Countries
Capability	Power	Water	Energy
Utilities	Maintenance	Generation	Quality
Capacity	Countries	Water	Energy
Operation	Company	Portfolio	Energy
Operation	Power	Assets	Energy
Generation	Power	Power	Energy
Facilities	Technology	Activity	Household
Provision	Businesses	Generation	Energy
Engineering	Buildings	Assets	Activity
Water	Range	businesses	Company
Resources	Power	Company	Provision
Facilities	Energy	Activities	Technology
Company	Power	Company	Energy
Energy	Power	Engineering	Energy
Provision	Performance	Company	Energy
Engineering	Power	Business	Technologies
Project	Energy	Engineering	Businesses
Company	Installation	Division	Offer
Energy	Power	Comprises	Household
Resources	Power	Generation	Energy
Energy	Energy	Power	Power
Power	Maintenance	Division	Provide
Division	Engineering	Operation	Company
Engineering	Operations	Generation	Power

Engineering	Company	Power	Household
Businesses	Power	Wind	Energy
Operation	Technology	Production	Technology
Power	Performance	Comprises	Energy
Facilities	Years	Production	Investment
Provision	Power	Technology	Business
Engineering	Energy	Company	Plan
Water	Production	Technology	Businesses
Resources	Sun	Energy	Energy
Facilities	Sun	Resources	Range
Company	systems	Water	Home
Energy	development	Water	Energy
Resources	plant	Business	Years
Energy	plant	Engineering	Energy
Division	january	Engineering	Energy
Engineering	november	Engineering	Energy
Engineering	history	Engineering	limited
Division	craftsmanship	Energy	works
Energy	fact	Water	increase
Division	cell	Years	people
Energy	percent	Generation	date
Resources	conversion	Water	million
Energy	cell	Portfolio	people
Facilities	system	Assets	million
Company	data	Power	million
Capacity	utility	Activity	support
Engineering	services	Operation	development
Division	equipment	Wind	limited
Engineering	small	Energy	hardware
Year	market	Generation	equipment
Water	dealer	Assets	performing
Design	network	Production	necessary
Power	direct	Businesses	light
Generation	europe	Technology	system
Facilities	systems	Company	development
Water	solar	spain	work
Engineering	public	sector	meet
Division	solar	main	help
Design	projects	station	advice
Energy	long	transmission	impetus
Division	term	concession	date
Energy	electricity	waste	million
Businesses	solar	treatment	people

Energy	better	development	result
Range	solar	may	million
Generation	real	environment	use
Power	world	electricity	million
Operation	world	sea	support
Facilities	record	market	development
Company	today	experience	world
Energy	high	gas	africa
Energy	world	large	alternative
Energy	record	scale	developing
Supply	level	transmission	access
Capability	average	concession	modern
Company	percent	concession	access
Energy	real	long	developing
Resources	world	term	alternative
Plan	long	pay	dealing
Resources	term	purchase	principle
Generation	high	transmission	green
Energy	percent	demand	develop
Capacity	plus	focus	alternative
Operation	low	high	subsidiaries
Grid	uptime	development	diode
Grid	corporation	leadership	lighting
Power	segment	position	highway
development	segment	practice	lighting
ownership	segment	type	subsidiary
distribution	procurement	infrastructure	research
services	segment	cogeneration	alternative
services	third	thermal	intellectual
include	party	subsidiary	specific
funding	ground	sustainability	develop
development	perform	biomass	access
management	financing	traditional	
services	efficiency	thermal	
cover	efficiency	gas	
development	efficiency	hybrid	
ownership	fast	type	
public	minimize	cogeneration	
stock	minimize	risk	
exchange	perform	component	
subsidiary		panorama	

owner		plant	
operator		finding	
columbia			
history			
services			
control			
use			
planning			
meet			
utility			
high			
remote			
program			
assist			
remote			
financing			
funding			
access			
capital			
markets			
term			
formation			
capital			
major			
utility			
selection			
near			
columbia			
connection			
north			
michigan			
line			
end			
interest			
sigma			
sigma			
trading			
sigma			
developer			
sigma			
competency			
advantage			
technical			
assistance			

develop			
diesil			
thermal			
origin			
initiative			
community			
identify			
technical			
expertise			
stability			
technical			
expertise			
growing			
trend			
future			
using			
mix			
independent			
develop			
vancouver			

9. The extracted instances from queries when WUP >= 0.7

QI 1	QI 2	QI 3	QI 4	QI 5	QI 6	QI 7
company	generation	wind	wind	engineering	wind	production
power	current	energy	energy	buildings	energy	wind
energy	energy	solarize	energy	energy	energy	energy
power	facilities	countries	wave	household	project	power
wind	range	generation	generation	home	china	power
power	energy	sun	wave	power	power	supply
design	power	power	wind	energy	energy	require
operations	capabilities	buinesses	type	resources	company	average
years	high	supply	fuel	wind	range	amount
require	high	year	speed	water	value	process
high	increase	company	type	sun	cost	type
amount	area	require	fourth	high	type	seventh
area	second	third	query	specification	average	query
type	limit	query	need	nature	consumption	liquid
matter	limit	impact	user	fifth	area	user
average	float	middle		query	sixth	
type	query	east			query	
experiment	specific	five				
kind	given					

affect	working					
	down					

QI 8	QI 9	QI 10	QI 11	QI 12	QI 13	QI 14	QI 15
engineering	engineering	company	energy	current	energy	businesses	company
wind	energy	water	water	wind	wind	company	energy
energy	wind	resources	power	energy	generation	wind	power
field	company	energy	energy	energy	energy	energy	wind
energy	might	air	company	company	activities	energy	energy
power	offer	current	require	water	businesses	power	energy
value	company	power	type	power	company	energy	company
business	require	energy	average	energy	type	wind	offer
land	type	production	consumption	buildings	area	quality	buildings
eighth	plants	company	float	type	support	company	type
query	services	average	area	average	type	type	average
experiment	major	type	support	consumption	thirteenth	type	consumption
total	nineth	area	type	type	query	close	require
experiment	query	require	eleventh	twelveth	user	area	area
		support	query	query		type	distance
		type	query	query		fourteenth	type
		tenth	user			query	fifteenth
		query				suggest	query
		consideration				user	query
						query	user

QI 16	QI 17	QI 18	QI 19	QI 20	QI 21	QI 22
company	company	wind	energy	china	company	wind
design	investment	power	water	power	wind	energy
wind	wind	energy	energy	water	energy	air
generation	company	products	power	energy	energy	current
resources	technologies	provide	technologies	energy	production	energy
energy	products	wind	products	supply	power	power
power	services	energy	businesses	capacity	company	energy
businesses	support	countries	type	water	technologies	facilities
average	cover	power	average	design	place	company
consumption	area	energy	limit	region	company	support
sixteenth	type	company	set	type	type	type
query	propose	energy	type	plus	amount	consumption
	seventeenth	manufacturer	nineteenth	area	support	type
	query	type	query	support	equipment	first
	superior	average	cross	type	near	experiment

		type	set	twentieth	first	higher
		eighteenth	time	query	working	cosideration
		query	user	large		
		experiment		scale		
				time		
				query		
				user		

QI 23	QI 24	QI 25	QI 26	QI 27	QI 28	QI 29	QI 30
water	company	company	energy	energy	wind	power	company
energy	wind	wind	supply	resources	energy	company	power
energy	power	water	energy	wind	generation	power	resources
facilities	wind	power	power	power	wind	energy	power
water	energy	energy	wind	production	china	company	energy
energy	energy	power	resources	solarize	company	station	energy
technologies	supply	energy	wind	range	type	infrastructure	technologies
power	company	group	design	heat	require	transmission	energy
power	years	type	wave	electricity	fuel	average	range
type	type	estimate	company	footprint	vehicle	consumption	development
set	area	type	type	environment	type	area	alternative
electricity	support	area	area	electricity	eighth	support	increase
use	type	support	support	seventh	query	type	type
include	fourth	fifth	type	query	apply	nineth	thirtieth
plants	query	query	sixth	case	user	query	query
plants	query	user	query	user		electric	intellectual
plants			user			query	stable
average						user	
consumption							
type							
flash							
suggest							
query							

10. Semantic Matching between the above instances WUP >= 0.7

	P1	Human	**Error**	P2	Human	**Error**	P3	Human	**Error**	P4	Human	**Error**
Q 1	40%	70%	**30%**	35%	60%	**25%**	15%	55%	**40%**	30%	65%	**35%**
Q 2	24%	60%	**36%**	33%	70%	**37%**	24%	50%	**26%**	19%	55%	**36%**

Q 3	28%	50%	**22%**	33%	65%	**32%**	11%	40%	**29%**	33%	40%	**7%**
Q 4	20%	55%	**35%**	33%	60%	**27%**	20%	60%	**40%**	27%	50%	**23%**
Q 5	19%	40%	**21%**	38%	55%	**17%**	31%	50%	**19%**	38%	65%	**27%**
Q 6	18%	50%	**32%**	41%	50%	**9%**	24%	55%	**31%**	53%	65%	**12%**
Q 7	25%	45%	**20%**	38%	75%	**37%**	6%	35%	**29%**	38%	75%	**37%**
Q 8	29%	40%	**11%**	43%	60%	**17%**	21%	50%	**29%**	57%	70%	**13%**
Q 9	29%	50%	**21%**	29%	65%	**36%**	29%	65%	**36%**	57%	70%	**13%**
Q 10	26%	65%	**39%**	37%	60%	**23%**	32%	50%	**18%**	42%	55%	**13%**
Q 11	11%	60%	**49%**	24%	75%	**51%**	24%	70%	**46%**	35%	65%	**30%**
Q 12	25%	65%	**40%**	38%	60%	**22%**	44%	70%	**26%**	38%	55%	**17%**
Q 13	21%	60%	**39%**	43%	80%	**37%**	21%	55%	**34%**	36%	65%	**29%**
Q 14	30%	60%	**30%**	35%	60%	**25%**	25%	65%	**40%**	45%	60%	**15%**
Q 15	20%	55%	**35%**	35%	85%	**50%**	30%	60%	**30%**	45%	70%	**25%**
Q 16	45%	60%	**15%**	55%	70%	**15%**	27%	55%	**28%**	45%	60%	**15%**
Q 17	25%	50%	**25%**	25%	65%	**40%**	13%	45%	**32%**	50%	60%	**10%**
Q 18	26%	60%	**34%**	47%	70%	**23%**	32%	65%	**33%**	53%	70%	**17%**
Q 19	11%	50%	**39%**	22%	60%	**38%**	28%	55%	**27%**	22%	50%	**28%**
Q 20	14%	50%	**36%**	18%	55%	**37%**	27%	50%	**23%**	36%	60%	**24%**
Q 21	41%	75%	**34%**	53%	85%	**32%**	18%	40%	**22%**	53%	60%	**7%**
Q 22	24%	60%	**36%**	41%	70%	**29%**	29%	65%	**36%**	41%	55%	**14%**
Q 23	20%	50%	**30%**	25%	75%	**50%**	20%	65%	**45%**	25%	65%	**40%**
Q 24	38%	60%	**22%**	44%	65%	**21%**	25%	60%	**35%**	63%	75%	**12%**
Q 25	25%	65%	**40%**	38%	75%	**37%**	25%	65%	**40%**	50%	70%	**20%**
Q 26	29%	60%	**31%**	35%	65%	**30%**	24%	55%	**31%**	59%	65%	**6%**
Q 27	31%	55%	**24%**	31%	60%	**29%**	19%	35%	**16%**	25%	65%	**40%**
Q 28	27%	60%	**33%**	33%	75%	**42%**	13%	50%	**37%**	33%	70%	**37%**
Q 29	22%	70%	**48%**	28%	85%	**57%**	17%	75%	**58%**	39%	80%	**41%**
Q 30	35%	50%	**15%**	53%	70%	**17%**	24%	60%	**36%**	53%	65%	**12%**

P5	Huamn	**Error**	P6	Human	**Error**	P7	Human	**Error**	P8	Human	**Error**	P9	Human	**Error**
35%	55%	**20%**	30%	50%	**20%**	45%	65%	**20%**	40%	70%	**30%**	30%	60%	**30%**
29%	50%	**21%**	43%	75%	**32%**	29%	40%	**11%**	29%	50%	**21%**	29%	50%	**21%**
33%	55%	**22%**	33%	60%	**27%**	39%	60%	**21%**	33%	60%	**27%**	28%	50%	**22%**
27%	55%	**28%**	27%	50%	**23%**	13%	20%	**7%**	40%	90%	**50%**	13%	40%	**27%**
44%	65%	**21%**	44%	45%	**1%**	44%	50%	**6%**	50%	65%	**15%**	31%	55%	**24%**
47%	65%	**18%**	41%	60%	**19%**	47%	65%	**18%**	35%	55%	**20%**	35%	60%	**25%**
19%	55%	**36%**	31%	65%	**34%**	31%	55%	**24%**	31%	70%	**39%**	19%	45%	**26%**

50%	75%	**25%**	29%	40%	**11%**	36%	60%	**24%**	43%	60%	**17%**	29%	60%	**31%**
36%	65%	**29%**	43%	60%	**17%**	36%	55%	**19%**	36%	65%	**29%**	29%	55%	**26%**
32%	70%	**38%**	37%	75%	**38%**	37%	50%	**13%**	37%	60%	**23%**	32%	50%	**18%**
29%	80%	**51%**	29%	85%	**56%**	29%	75%	**46%**	29%	90%	**61%**	29%	75%	**46%**
38%	75%	**37%**	38%	60%	**22%**	44%	60%	**16%**	44%	70%	**26%**	31%	60%	**29%**
29%	70%	**41%**	36%	65%	**29%**	29%	50%	**21%**	50%	90%	**40%**	36%	55%	**19%**
40%	75%	**35%**	35%	65%	**30%**	35%	60%	**25%**	35%	75%	**40%**	40%	85%	**45%**
30%	75%	**45%**	30%	60%	**30%**	40%	70%	**30%**	30%	70%	**40%**	35%	70%	**35%**
36%	70%	**34%**	64%	75%	**11%**	27%	40%	**13%**	55%	80%	**25%**	27%	50%	**23%**
50%	80%	**30%**	31%	50%	**19%**	31%	55%	**24%**	25%	60%	**35%**	44%	75%	**31%**
42%	75%	**33%**	37%	60%	**23%**	53%	70%	**17%**	47%	80%	**33%**	53%	70%	**17%**
44%	85%	**41%**	28%	65%	**37%**	33%	60%	**27%**	28%	65%	**37%**	33%	80%	**47%**
40%	80%	**40%**	36%	80%	**44%**	14%	40%	**26%**	18%	50%	**32%**	27%	40%	**13%**
41%	65%	**24%**	41%	55%	**14%**	47%	75%	**28%**	47%	70%	**23%**	47%	60%	**13%**
29%	65%	**36%**	35%	65%	**30%**	29%	60%	**31%**	29%	60%	**31%**	35%	45%	**10%**
42%	75%	**33%**	42%	75%	**33%**	29%	50%	**21%**	20%	50%	**30%**	29%	55%	**26%**
38%	60%	**22%**	38%	60%	**22%**	38%	65%	**27%**	50%	80%	**30%**	44%	60%	**16%**
38%	75%	**37%**	38%	80%	**42%**	31%	75%	**44%**	44%	85%	**41%**	38%	65%	**27%**
29%	60%	**31%**	35%	55%	**20%**	24%	50%	**26%**	41%	65%	**24%**	29%	30%	**1%**
31%	50%	**19%**	25%	50%	**25%**	38%	60%	**22%**	50%	75%	**25%**	19%	35%	**16%**
27%	65%	**38%**	20%	40%	**20%**	13%	30%	**17%**	33%	75%	**42%**	13%	30%	**17%**
28%	85%	**57%**	28%	80%	**52%**	33%	85%	**52%**	39%	90%	**51%**	33%	80%	**47%**
53%	85%	**32%**	53%	85%	**32%**	35%	60%	**25%**	53%	80%	**27%**	59%	75%	**16%**

11. The extracted instances form providers when WUP >= 0.6

Provider1	**Provider2**	**Provider3**	**Provider4**	**Provider5**
Generation	Wind	Water	Wind	China
Design	Power	Water	Wind	Industry
Performance	Wind	Design	Offer	Power
Wind	Production	Installation	Business	Power
Design	Wind	Assets	Wind	Power
Operations	Generation	Water	Power	Capacity
Wind	Wind	Installation	Wind	Water
Production	Generation	Facility	Wind	Businesses
Wind	Power	Water	Wind	Minutes
Capabilities	Transaction	Energy	Wind	Power
Wind	Transactions	Water	Power	Power
Wind	Business	Energy	Business	Investment
Wind	Investment	Design	Technology	Home
Operations	Wind	Water	Business	Performance
Operations	Generation	Water	Business	Products
Wind	Facilities	Water	Business	Technologies

Design	Wind	Water	Technology	Products
Wind	Wind	Buildings	Business	Design
Design	Power	Design	Power	Products
Wind	Activities	Installation	Power	Performance
Capacity	Power	Current	Production	Quality
Generation	Power	Company	Energy	Power
Power	Wind	Assets	Company	Products
Power	Energy	Engineering	Wind	China
Wind	Wind	Energy	Energy	China
Power	Energy	Water	Wind	Industry
Power	Energy	Water	Technology	Industry
Wind	Production	Facility	Products	Company
Power	Company	Provide	Offer	Power
Wind	Company	Design	Energy	Products
Current	Wind	Provide	Business	Company
Wind	Energy	installation	Maintenance	Products
Generation	Generation	manufacture	Wind	Power
Capacity	Wind	source	Power	Energy
Technology	Energy	systems	Value	Capacity
Design	Generation	december	Maintenance	Food
Company	Life	source	Wind	Water
Business	Power	systems	Capacity	Energy
Wind	Value	distribution	Wind	Businesses
Engineering	Industry	pacific	Wind	Energy
Design	Ability	heating	Company	Sake
Production	Business	cooling	Wind	Energy
Engineering	Investment	systems	Power	Energy
Wind	Lead	coductivity	Industry	Energy
Operation	Investment	assistance	Company	Sunlight
Technologies	Wind	limit	Business	Power
Company	Generation	draw	Technology	Energy
Operations	Facilities	start	Business	Businesses
Company	Facilities	january	Business	Power
Wind	Wind	footprint	Business	Grid
Capabilities	Generation	time	China	Year
Wind	Facility	use	Business	Energy
Engineering	Energy	free	Technology	Years
Wind	Wind	heating	Business	Investment
Operations	Energy	cooling	Engineering	Business
Quality	Energy	yard	Engineering	Energy
Design	Company	source	Company	Energy
Wind	Value	heat	China	Home
Installation	Power	pump	Power	Value
Performance	Resources	system	Energy	Energy

Design	Business	state	Resources	Quality
Installation	play	art	Energy	Performance
Lift	asset	equipment	Energy	Products
Generation	Life	people	Energy	Technologies
Grid	provider	check	Power	Quality
Years	term	start	Wind	Products
electricity	cash	earth	Energy	Project
close	appreciate	testing	Quality	Engineering
increase	time	devices	Life	Design
assembly	strategy	inventory	Energy	Project
manufacturing	combine	found	Supply	Industry
development	Industry	back	Energy	Products
maintenance	expertise	save	Food	Performance
include	Ability	costs	Production	Company
control	capital	check	Energy	Quality
system	order	subsidary	superior	Field
development	secure	inrush	cost	Field
manufacture	benefit	electric	services	limited
sale	development	binary	works	limited
north	strategy	carbon	close	manufacturer
south	platform	same	partnership	distribution
america	near		core	network
manufacturing	march		development	europe
development	interest		manufacturing	north
maintenance	western		sale	america
management	may		cover	asia
weight	strategy		chain	covering
weight	high		site	stock
light	capital		service	exchange
manufacture	allocation		market	manufacture
transport	build		leading	system
maintenance	buy		position	integration
transport	understanding		day	leader
connection	public		demand	services
standard	safety		systems	standard
size	values		manufacturing	output
connection	partnership		service	markets
five	open		area	number
farm	communication		support	germany
segment	manage		asia	spain
segment	ways		europe	italy
africa	ensure		mediterranean	system
main	development		europe	december
variety	appraoch		manufacturing	manufacturing

farm	play		area	rapid
low	primary		manufacturing	development
play	vehicle		assembly	use
primary	one		service	world
vehicle	arizona		area	economy
growing	california		platform	dioxide
limited	stable		management	climate
light	over		services	change
	focus		december	balance
	california		network	even
	subsidary		germany	fuel
	focus		italy	rise
	accept		spain	sake
	natural		sweden	environment
	three		norway	finding
	one		use	today
	commit		pressure	world
	three		change	source
	equity		way	world
	growing		produce	reach
	buy		produce	germany
	fundamental		source	italy
	accountability		catalyst	reach
	manage		better	cleaner
	two		stable	family
			essential	electricity
			part	costs
			socirty	increase
			major	top
			challenge	high
			world	world
			today	setting
			security	third
			climate	high
			change	test
			food	time
			access	financing
			strengthening	track
			equity	step
			independence	module
			total	deployment
			bringing	silicon
			par	pioneer
			gas	shaper

			growing	european
			denmark	plant
			active	phoenix
			four	CO2
			six	natural
			central	crisis
			denmark	worsening
			sustainability	pollution
			growing	crisis
			planet	modern
			great	inequality
			founding	facing
			infrastructure	ready
			opportunity	Safe
			sustainability	Like
			strengthening	California
			poverty	Possible
			oil	Reduce
			four	Bill
			finance	Indendence
			sales	Rising
			finance	Record
			sales	Party
			six	Testing
			pacific	Stand
			reducing	Track
				Silicon
				Step
				Become
				Annual
				Fossil
				Carbon
				Nothing
				Ecosystem
				Become
				Enough
				Entire
				Clean
				Two
				Three
				Parity
				Raly

Provier6	Provider7	Provider8	Provider9	Provider10
Product	Operation	Business	Company	Businesses
Mexico	Generation	Products	Engineering	Countries
Countries	Water	construction	Company	Quality
Company	Resources	Power	Business	Home
Company	Power	Power	Construction	Technologies
Food	businesses	Power	Division	Products
Addition	Power	Company	Comprises	Activity
Products	Facilities	Technology	Execution	Company
Home	Water	Maintenance	Generation	Investments
Countries	Resources	businesses	Division	Businesses
Company	Engineering	Buildings	Operation	Offer
Mexico	Resources	Range	Generation	Methods
Company	Capacity	Performance	Power	Combination
Company	Engineering	Installation	Production	Power
Product	Design	Power	Comprises	Power
Food	Power	Business	Activities	Technology
development	Water	Power	Production	Range
stage	Division	Power	Technology	Products
distribute	Design	Business	Water	Business
free	Range	Power	Business	Initiatives
spill	Utilities	Project	Construction	Energy
free	Demonstration	Project	Engineering	Business
cone	Grid	Power	Activities	Countries
cone	Capability	Power	Water	Energy
cone	Utilities	Maintenance	Generation	Quality
markrt	Capacity	Countries	Water	Energy
cone	Operation	Company	Portfolio	Energy
sell	Operation	Power	Assets	Energy
use	Generation	Power	Power	Energy
world	Facilities	Technology	Activity	Household
cone	Provision	Businesses	Generation	Energy
canada	Engineering	Buildings	Assets	Activity
president	Water	Range	businesses	Company
cheese	Resources	Power	Company	Provision
today	Facilities	Energy	Activities	Technology
cone	Company	Power	Company	Energy
number	Energy	Power	Engineering	Energy
variety	Provision	Performance	Company	Energy
breakfast	Engineering	Power	Business	Technologies
light	Project	Energy	Engineering	Businesses
desert	Company	Installation	Division	Offer
free	Energy	Power	Comprises	Household
bottom	Resources	Power	Generation	Energy

sunrise	Energy	Energy	Power	Power
drip	Power	Maintenance	Division	Provide
dough	Division	Engineering	Operation	Company
fast	Engineering	Operations	Generation	Power
catering	Engineering	Company	Power	Household
chief	Businesses	Power	Wind	Energy
officer	Operation	Technology	Production	Technology
milan	Power	Performance	Comprises	Energy
chief	Facilities	Years	Production	Investment
dough	Provision	Power	Technology	Business
proprietary	Engineering	Energy	Company	Plan
sauce	Water	Production	Technology	Businesses
snake	Resources	Sun	Energy	Energy
par	Facilities	Sun	Resources	Range
drip	Company	systems	Water	Home
bakery	Energy	development	Water	Energy
brooklyn	Resources	plant	Business	Years
	Energy	plant	Engineering	Energy
	Division	january	Engineering	Energy
	Engineering	november	Engineering	Energy
	Engineering	history	Engineering	limited
	Division	craftsmanship	Energy	works
	Energy	fact	Water	increase
	Division	cell	Years	people
	Energy	percent	Generation	date
	Resources	conversion	Water	million
	Energy	cell	Portfolio	people
	Facilities	system	Assets	million
	Company	data	Power	million
	Capacity	utility	Activity	support
	Engineering	services	Operation	development
	Division	equipment	Wind	limited
	Engineering	small	Energy	hardware
	Year	market	Generation	equipment
	Water	dealer	Assets	performing
	Design	network	Production	necessary
	Power	direct	Businesses	light
	Generation	europe	Technology	system
	Facilities	systems	Company	development
	Water	solar	spain	work
	Engineering	public	sector	meet
	Division	solar	main	help
	Design	projects	station	advice
	Energy	long	transmission	impetus

	Division	term	concession	date
	Energy	electricity	waste	million
	Businesses	solar	treatment	people
	Energy	better	development	result
	Range	solar	may	million
	Generation	real	environment	use
	Power	world	electricity	million
	Operation	world	sea	support
	Facilities	record	market	development
	Company	today	experience	world
	Energy	high	gas	africa
	Energy	world	large	alternative
	Energy	record	scale	developing
	Supply	level	transmission	access
	Capability	average	concession	modern
	Company	percent	concession	access
	Energy	real	long	developing
	Resources	world	term	alternative
	Plan	long	pay	dealing
	Resources	term	purchase	principle
	Generation	high	transmission	green
	Energy	percent	demand	develop
	Capacity	plus	focus	alternative
	Operation	low	high	subsidiaries
	Grid	uptime	development	diode
	Grid	corporation	leadership	lighting
	Power	segment	position	highway
	development	segment	practice	lighting
	ownership	segment	type	subsidiary
	distribution	procurement	infrastructure	research
	services	segment	cogeneration	alternative
	services	third	thermal	intellectual
	include	party	subsidiary	specific
	funding	ground	sustainability	develop
	development	perform	biomass	access
	management	financing	traditional	local
	services	efficiency	thermal	improve
	cover	efficiency	gas	clean
	development	efficiency	hybrid	local
	ownership	fast	type	clean
	public	minimize	cogeneration	bulb
	stock	minimize	risk	four
	exchange	perform	component	clean
	subsidiary	two	panorama	five

	owner	rooftop	plant	local
	operator	two	finding	caribbean
	columbia	ten	three	
	history	deliver	turnkey	
	services		electric	
	control		three	
	use		turnkey	
	planning		take	
	meet		pay	
	utility		electric	
	high			
	remote			
	program			
	assist			
	remote			
	financing			
	funding			
	access			
	capital			
	markets			
	term			
	formation			
	capital			
	major			
	utility			
	selection			
	near			
	columbia			
	connection			
	north			
	michigan			
	line			
	end			
	interest			
	sigma			
	sigma			
	trading			
	sigma			
	developer			
	sigma			
	competency			
	advantage			
	technical			
	assistance			

	develop			
	diesil			
	thermal			
	origin			
	initiative			
	community			
	identify			
	technical			
	expertise			
	stability			
	technical			
	expertise			
	growing			
	trend			
	future			
	using			
	mix			
	independent			
	develop			
	vancouver			
	two			
	electric			
	trading			
	toronto			
	reputation			
	federal			
	financing			
	access			
	rupert			
	british			
	three			

11. The extracted instances from queries when WUP >= 0.6

QI 1	QI 2	QI 3	QI 4	QI 5	QI 6	QI 7	QI 8
company	generation	wind	wind	engineering	wind	production	engineering
power	current	energy	energy	buildings	energy	wind	wind
energy	energy	solarize	energy	energy	energy	energy	energy
power	facilities	countries	wave	household	project	power	field
wind	range	generation	generation	home	china	power	energy
power	energy	sun	wave	power	power	supply	power
design	power	power	wind	energy	energy	require	value

operations	capabilities	buinesses	type	resources	company	average	business
years	high	supply	fuel	wind	range	amount	land
require	high	year	speed	water	value	process	eighth
high	increase	company	type	sun	cost	type	query
amount	area	require	fourth	high	type	seventh	experiment
area	second	third	query	specification	average	query	total
type	limit	query	need	nature	consumption	liquid	experiment
matter	limit	impact	user	fifth	area	user	
average	float	middle		query	sixth	factory	
type	query	east			query		
experiment	specific	five					
Kind	given	car					
Affect	working	green					
Consume	down	positive					
	factory						
	manager						

QI 9	QI 10	QI 11	QI 12	QI 13	QI 14	QI 15	QI 16
engineering	company	energy	current	energy	businesses	company	company
energy	water	water	wind	wind	company	energy	design
wind	resources	power	energy	generation	wind	power	wind
company	energy	energy	energy	energy	energy	wind	generation
might	air	company	company	activities	energy	energy	resources
offer	current	require	water	businesses	power	energy	energy
company	power	type	power	company	energy	company	power
require	energy	average	energy	type	wind	offer	businesses
type	production	consumption	buildings	area	quality	buildings	average
plants	company	float	type	support	company	type	consumption
services	average	area	average	type	type	average	sixteenth
major	type	support	consumption	thirteenth	type	consumption	query
nineth	area	type	type	query	close	require	
query	require	eleventh	twelveth	user	area	area	
	support	query	query		type	distance	
	type	query	query		fourteeenth	type	
	tenth	user			query	fifteenth	
	query				suggest	query	
	consideration				user	query	
					query	user	

QI 17	QI 18	QI 19	QI 20	QI 21	QI 22	QI 23
company	wind	energy	china	company	wind	water

investment	power	water	power	wind	energy	energy
wind	energy	energy	water	energy	air	energy
company	products	power	energy	energy	current	facilities
technologies	provide	technologies	energy	production	energy	water
products	wind	products	supply	power	power	energy
services	energy	businesses	capacity	company	energy	technologies
support	countries	type	water	technologies	facilities	power
cover	power	average	design	place	company	power
area	energy	limit	region	company	support	type
type	company	set	type	type	type	set
propose	energy	type	plus	amount	consumption	electricity
seventeenth	manufacturer	nineteenth	area	support	type	use
query	type	query	support	equipment	first	include
superior	average	cross	type	near	experiment	plants
	type	set	twentieth	first	higher	plants
	eighteenth	time	query	working	cosideration	plants
	query	user	large			average
	experiment		scale			consumption
			time			type
			query			flash
			user			suggest
						query
						dry
						steam
						steam
						binary
						cycle

QI 24	QI 25	QI 26	QI 27	QI 28	QI 29	QI 30
company	company	energy	energy	wind	power	company
wind	wind	supply	resources	energy	company	power
power	water	energy	wind	generation	power	resources
wind	power	power	power	wind	energy	power
energy	energy	wind	production	china	company	energy
energy	power	resources	solarize	company	station	energy
supply	energy	wind	range	type	infrastructure	technologies
company	group	design	heat	require	transmission	energy
years	type	wave	electricity	fuel	average	range
type	estimate	company	footprint	vehicle	consumption	development
area	type	type	environment	type	area	alternative
support	area	area	electricity	eighth	support	increase
type	support	support	seventh	query	type	type

fourth	fifth	type	query	apply	nineth	thirtieth
query	query	sixth	case	user	query	query
query	user	query	user		electric	intellectual
commit		user	reduce		query	stable
		factory			user	

12. Semantic Matching between instances when WUP >= 0.6

	P1	Human	**Error**	P2	Human	**Error**	P3	Human	**Error**	P4	Human	**Error**
Q 1	38%	70%	**32%**	33%	60%	**27%**	14%	55%	**41%**	29%	65%	**36%**
Q 2	22%	60%	**38%**	30%	70%	**40%**	22%	50%	**28%**	17%	55%	**38%**
Q 3	24%	50%	**26%**	29%	65%	**36%**	9%	40%	**31%**	29%	40%	**11%**
Q 4	20%	55%	**35%**	33%	60%	**27%**	20%	60%	**40%**	27%	50%	**23%**
Q 5	19%	40%	**21%**	38%	55%	**17%**	31%	50%	**19%**	38%	65%	**27%**
Q 6	18%	50%	**32%**	41%	50%	**9%**	24%	55%	**31%**	53%	65%	**12%**
Q 7	24%	45%	**21%**	35%	75%	**40%**	5%	35%	**30%**	35%	75%	**40%**
Q 8	29%	40%	**11%**	43%	60%	**17%**	21%	50%	**29%**	57%	70%	**13%**
Q 9	29%	50%	**21%**	29%	65%	**36%**	29%	65%	**36%**	57%	70%	**13%**
Q 10	26%	65%	**39%**	37%	60%	**23%**	32%	50%	**18%**	42%	55%	**13%**
Q 11	11%	60%	**49%**	24%	75%	**51%**	24%	70%	**46%**	35%	65%	**30%**
Q 12	25%	65%	**40%**	38%	60%	**22%**	44%	70%	**26%**	38%	55%	**17%**
Q 13	21%	60%	**39%**	43%	80%	**37%**	21%	55%	**34%**	36%	65%	**29%**
Q 14	30%	60%	**30%**	35%	60%	**25%**	25%	65%	**40%**	45%	60%	**15%**
Q 15	20%	55%	**35%**	35%	85%	**50%**	30%	60%	**30%**	45%	70%	**25%**
Q 16	45%	60%	**15%**	55%	70%	**15%**	27%	55%	**28%**	45%	60%	**15%**
Q 17	25%	50%	**25%**	25%	65%	**40%**	13%	45%	**32%**	50%	60%	**10%**
Q 18	26%	60%	**34%**	47%	70%	**23%**	32%	65%	**33%**	53%	70%	**17%**
Q 19	11%	50%	**39%**	22%	60%	**38%**	28%	55%	**27%**	22%	50%	**28%**
Q 20	14%	50%	**36%**	18%	55%	**37%**	27%	50%	**23%**	36%	60%	**24%**
Q 21	41%	75%	**34%**	53%	85%	**32%**	18%	40%	**22%**	53%	60%	**7%**
Q 22	24%	60%	**36%**	41%	70%	**29%**	29%	65%	**36%**	41%	55%	**14%**
Q 23	17%	50%	**33%**	20%	75%	**55%**	17%	65%	**48%**	20%	65%	**45%**
Q 24	35%	60%	**25%**	41%	65%	**24%**	35%	60%	**25%**	59%	75%	**16%**
Q 25	25%	65%	**40%**	38%	75%	**37%**	25%	65%	**40%**	50%	70%	**20%**
Q 26	28%	60%	**32%**	33%	65%	**32%**	22%	55%	**33%**	56%	65%	**9%**
Q 27	29%	55%	**26%**	29%	60%	**31%**	18%	35%	**17%**	24%	65%	**41%**

Q 28	27%	60%	**33%**	33%	75%	**42%**	13%	50%	**37%**	33%	70%	**37%**
Q 29	22%	70%	**48%**	28%	85%	**57%**	17%	75%	**58%**	39%	80%	**41%**
Q 30	35%	50%	**15%**	53%	70%	**17%**	24%	60%	**36%**	53%	65%	**12%**

P5	Huamn	**Error**	P6	Human	**Error**	P7	Human	**Error**	P8	Human	**Error**	P9	Human	**Error**
33%	55%	**22%**	29%	50%	**21%**	43%	65%	**22%**	38%	70%	**32%**	29%	60%	**31%**
26%	50%	**24%**	39%	75%	**36%**	26%	40%	**14%**	26%	50%	**24%**	26%	50%	**24%**
29%	55%	**26%**	29%	60%	**31%**	33%	60%	**27%**	29%	60%	**31%**	29%	50%	**21%**
27%	55%	**28%**	27%	50%	**23%**	13%	20%	**7%**	40%	90%	**50%**	13%	40%	**27%**
44%	65%	**21%**	44%	45%	**1%**	44%	50%	**6%**	50%	65%	**15%**	31%	55%	**24%**
47%	65%	**18%**	41%	60%	**19%**	47%	65%	**18%**	35%	55%	**20%**	35%	60%	**25%**
18%	55%	**37%**	29%	65%	**36%**	29%	55%	**26%**	29%	70%	**41%**	18%	45%	**27%**
50%	75%	**25%**	29%	40%	**11%**	36%	60%	**24%**	43%	60%	**17%**	29%	60%	**31%**
36%	65%	**29%**	43%	60%	**17%**	36%	55%	**19%**	36%	65%	**29%**	29%	55%	**26%**
32%	70%	**38%**	37%	75%	**38%**	37%	50%	**13%**	37%	60%	**23%**	32%	50%	**18%**
29%	80%	**51%**	29%	85%	**56%**	29%	75%	**46%**	29%	90%	**61%**	29%	75%	**46%**
38%	75%	**37%**	38%	60%	**22%**	44%	60%	**16%**	44%	70%	**26%**	31%	60%	**29%**
29%	70%	**41%**	36%	65%	**29%**	29%	50%	**21%**	50%	90%	**40%**	36%	55%	**19%**
40%	75%	**35%**	35%	65%	**30%**	35%	60%	**25%**	35%	75%	**40%**	40%	85%	**45%**
30%	75%	**45%**	30%	60%	**30%**	40%	70%	**30%**	30%	70%	**40%**	35%	70%	**35%**
36%	70%	**34%**	64%	75%	**11%**	27%	40%	**13%**	55%	80%	**25%**	27%	50%	**23%**
50%	80%	**30%**	31%	50%	**19%**	31%	55%	**24%**	25%	60%	**35%**	44%	75%	**31%**
42%	75%	**33%**	37%	60%	**23%**	53%	70%	**17%**	47%	80%	**33%**	53%	70%	**17%**
44%	85%	**41%**	28%	65%	**37%**	33%	60%	**27%**	28%	65%	**37%**	33%	80%	**47%**
40%	80%	**40%**	36%	80%	**44%**	14%	40%	**26%**	18%	50%	**32%**	27%	40%	**13%**
41%	65%	**24%**	41%	55%	**14%**	47%	75%	**28%**	47%	70%	**23%**	47%	60%	**13%**
29%	65%	**36%**	35%	65%	**30%**	29%	60%	**31%**	29%	60%	**31%**	35%	45%	**10%**
34%	75%	**41%**	34%	75%	**41%**	24%	50%	**26%**	17%	50%	**33%**	24%	55%	**31%**
35%	60%	**25%**	35%	60%	**25%**	35%	65%	**30%**	47%	80%	**33%**	41%	60%	**19%**
38%	75%	**37%**	38%	80%	**42%**	31%	75%	**44%**	44%	85%	**41%**	38%	65%	**27%**
28%	60%	**32%**	33%	55%	**22%**	22%	50%	**28%**	39%	65%	**26%**	28%	30%	**2%**
35%	50%	**15%**	24%	50%	**26%**	35%	60%	**25%**	47%	75%	**28%**	18%	35%	**17%**
27%	65%	**38%**	20%	40%	**20%**	13%	30%	**17%**	33%	75%	**42%**	13%	30%	**17%**
28%	85%	**57%**	28%	80%	**52%**	33%	85%	**52%**	39%	90%	**51%**	33%	80%	**47%**
53%	85%	**32%**	53%	85%	**32%**	35%	60%	**25%**	53%	80%	**27%**	59%	75%	**16%**

13. The extracted instances from providers when WUP >= 0.5

Provider1	**Provider2**	**Provider3**	**Provider4**	**Provider5**
Generation	Wind	Water	Wind	China

Design	Power	Water	Wind	Industry
Performance	Wind	Design	Offer	Power
Wind	Production	Installation	Business	Power
Design	Wind	Assets	Wind	Power
Operations	Generation	Water	Power	Capacity
Wind	Wind	Installation	Wind	Water
Production	Generation	Facility	Wind	Businesses
Wind	Power	Water	Wind	Minutes
Capabilities	Transaction	Energy	Wind	Power
Wind	Transactions	Water	Power	Power
Wind	Business	Energy	Business	Investment
Wind	Investment	Design	Technology	Home
Operations	Wind	Water	Business	Performance
Operations	Generation	Water	Business	Products
Wind	Facilities	Water	Business	Technologies
Design	Wind	Water	Technology	Products
Wind	Wind	Buildings	Business	Design
Design	Power	Design	Power	Products
Wind	Activities	Installation	Power	Performance
Capacity	Power	Current	Production	Quality
Generation	Power	Company	Energy	Power
Power	Wind	Assets	Company	Products
Power	Energy	Engineering	Wind	China
Wind	Wind	Energy	Energy	China
Power	Energy	Water	Wind	Industry
Power	Energy	Water	Technology	Industry
Wind	Production	Facility	Products	Company
Power	Company	Provide	Offer	Power
Wind	Company	Design	Energy	Products
Current	Wind	Provide	Business	Company
Wind	Energy	installation	Maintenance	Products
Generation	Generation	manufacture	Wind	Power
Capacity	Wind	source	Power	Energy
Technology	Energy	systems	Value	Capacity
Design	Generation	december	Maintenance	Food
Company	Life	source	Wind	Water
Business	Power	systems	Capacity	Energy
Wind	Value	distribution	Wind	Businesses
Engineering	Industry	pacific	Wind	Energy
Design	Ability	heating	Company	Sake
Production	Business	cooling	Wind	Energy
Engineering	Investment	systems	Power	Energy
Wind	Lead	coductivity	Industry	Energy
Operation	Investment	assistance	Company	Sunlight

Technologies	Wind	limit	Business	Power
Company	Generation	draw	Technology	Energy
Operations	Facilities	start	Business	Businesses
Company	Facilities	january	Business	Power
Wind	Wind	footprint	Business	Grid
Capabilities	Generation	time	China	Year
Wind	Facility	use	Business	Energy
Engineering	Energy	free	Technology	Years
Wind	Wind	heating	Business	Investment
Operations	Energy	cooling	Engineering	Business
Quality	Energy	yard	Engineering	Energy
Design	Company	source	Company	Energy
Wind	Value	heat	China	Home
Installation	Power	pump	Power	Value
Performance	Resources	system	Energy	Energy
Design	Business	state	Resources	Quality
Installation	play	art	Energy	Performance
Lift	asset	equipment	Energy	Products
Generation	Life	people	Energy	Technologies
Grid	provider	check	Power	Quality
Years	term	start	Wind	Products
electricity	cash	earth	Energy	Project
close	appreciate	testing	Quality	Engineering
increase	time	devices	Life	Design
assembly	strategy	inventory	Energy	Project
manufacturing	combine	found	Supply	Industry
development	Industry	back	Energy	Products
maintenance	expertise	save	Food	Performance
include	Ability	costs	Production	Company
control	capital	check	Energy	Quality
system	order	subsidary	superior	Field
development	secure	inrush	cost	Field
manufacture	benefit	electric	services	Limited
sale	development	binary	works	Limited
north	strategy	carbon	close	Manufacturer
south	platform	same	partnership	Distribution
america	near		core	Network
manufacturing	march		development	Europe
development	interest		manufacturing	North
maintenance	western		sale	America
management	may		cover	Asia
weight	strategy		chain	Covering
weight	high		site	Stock
light	capital		service	Exchange

manufacture	allocation		market	Manufacture
transport	build		leading	System
maintenance	buy		position	Integration
transport	understanding		day	Leader
connection	public		demand	Services
standard	safety		systems	Standard
size	values		manufacturing	Output
connection	partnership		service	Markets
five	open		area	Number
farm	communication		support	Germany
segment	manage		asia	Spain
segment	ways		europe	Italy
africa	ensure		mediterranean	System
main	development		europe	December
variety	appraoch		manufacturing	Manufacturing
farm	play		area	Rapid
low	primary		manufacturing	Development
play	vehicle		assembly	Use
primary	one		service	World
vehicle	arizona		area	Economy
growing	california		platform	Dioxide
limited	stable		management	Climate
light	over		services	Change
acquire	focus		december	Balance
	california		network	Even
	subsidary		germany	Fuel
	focus		italy	Rise
	accept		spain	Sake
	natural		sweden	environment
	three		norway	Finding
	one		use	Today
	commit		pressure	World
	three		change	Source
	equity		way	World
	growing		produce	Reach
	buy		produce	Germany
	fundamental		source	Italy
	accountability		catalyst	Reach
	manage		better	Cleaner
	two		stable	Family
	acquire		essential	Electricity
	international		part	Costs
	electric		socirty	Increase
	binary		major	Top

	carbon		challenge	High
	own		world	World
			today	Setting
			security	Third
			climate	High
			change	Test
			food	Time
			access	Financing
			strengthening	Track
			equity	Step
			independence	Module
			total	Deployment
			bringing	Silicon
			par	Pioneer
			gas	Shaper
			growing	European
			denmark	Plant
			active	Phoenix
			four	CO2
			six	Natural
			central	Crisis
			denmark	Worsening
			sustainability	Pollution
			growing	Crisis
			planet	Modern
			great	Inequality
			founding	Facing
			infrastructure	ready
			opportunity	safe
			sustainability	like
			strengthening	california
			poverty	possible
			oil	reduce
			four	bill
			finance	indendence
			sales	rising
			finance	record
			sales	party
			six	testing
			pacific	stand
			reducing	track
			four	silicon
			sales	step
			six	become

				annual
				fossil
				carbon
				nothing
				ecosystem
				become
				enough
				entire
				clean
				two
				three
				parity
				raly

Provier6	**Provider7**	**Provider8**	**Provider9**	**Provider10**
Product	Operation	Business	Company	Businesses
Mexico	Generation	Products	Engineering	Countries
Countries	Water	construction	Company	Quality
Company	Resources	Power	Business	Home
Company	Power	Power	Construction	Technologies
Food	businesses	Power	Division	Products
Addition	Power	Company	Comprises	Activity
Products	Facilities	Technology	Execution	Company
Home	Water	Maintenance	Generation	Investments
Countries	Resources	businesses	Division	Businesses
Company	Engineering	Buildings	Operation	Offer
Mexico	Resources	Range	Generation	Methods
Company	Capacity	Performance	Power	Combination
Company	Engineering	Installation	Production	Power
Product	Design	Power	Comprises	Power
Food	Power	Business	Activities	Technology
development	Water	Power	Production	Range
stage	Division	Power	Technology	Products
distribute	Design	Business	Water	Business
free	Range	Power	Business	Initiatives
spill	Utilities	Project	Construction	Energy
free	Demonstration	Project	Engineering	Business
cone	Grid	Power	Activities	Countries
cone	Capability	Power	Water	Energy
cone	Utilities	Maintenance	Generation	Quality
markrt	Capacity	Countries	Water	Energy
cone	Operation	Company	Portfolio	Energy
sell	Operation	Power	Assets	Energy

use	Generation	Power	Power	Energy
world	Facilities	Technology	Activity	Household
cone	Provision	Businesses	Generation	Energy
canada	Engineering	Buildings	Assets	Activity
president	Water	Range	businesses	Company
cheese	Resources	Power	Company	Provision
today	Facilities	Energy	Activities	Technology
cone	Company	Power	Company	Energy
number	Energy	Power	Engineering	Energy
variety	Provision	Performance	Company	Energy
breakfast	Engineering	Power	Business	Technologies
light	Project	Energy	Engineering	Businesses
desert	Company	Installation	Division	Offer
free	Energy	Power	Comprises	Household
bottom	Resources	Power	Generation	Energy
sunrise	Energy	Energy	Power	Power
drip	Power	Maintenance	Division	Provide
dough	Division	Engineering	Operation	Company
fast	Engineering	Operations	Generation	Power
catering	Engineering	Company	Power	Household
chief	Businesses	Power	Wind	Energy
officer	Operation	Technology	Production	Technology
milan	Power	Performance	Comprises	Energy
chief	Facilities	Years	Production	Investment
dough	Provision	Power	Technology	Business
proprietary	Engineering	Energy	Company	Plan
sauce	Water	Production	Technology	Businesses
snake	Resources	Sun	Energy	Energy
par	Facilities	Sun	Resources	Range
drip	Company	systems	Water	Home
bakery	Energy	development	Water	Energy
brooklyn	Resources	plant	Business	Years
	Energy	plant	Engineering	Energy
	Division	january	Engineering	Energy
	Engineering	november	Engineering	Energy
	Engineering	history	Engineering	limited
	Division	craftsmanship	Energy	works
	Energy	fact	Water	increase
	Division	cell	Years	people
	Energy	percent	Generation	date
	Resources	conversion	Water	million
	Energy	cell	Portfolio	people
	Facilities	system	Assets	million
	Company	data	Power	million

	Capacity	utility	Activity	support
	Engineering	services	Operation	development
	Division	equipment	Wind	limited
	Engineering	small	Energy	hardware
	Year	market	Generation	equipment
	Water	dealer	Assets	performing
	Design	network	Production	necessary
	Power	direct	Businesses	light
	Generation	europe	Technology	system
	Facilities	systems	Company	development
	Water	solar	spain	work
	Engineering	public	sector	meet
	Division	solar	main	help
	Design	projects	station	advice
	Energy	long	transmission	impetus
	Division	term	concession	date
	Energy	electricity	waste	million
	Businesses	solar	treatment	people
	Energy	better	development	result
	Range	solar	may	million
	Generation	real	environment	use
	Power	world	electricity	million
	Operation	world	sea	support
	Facilities	record	market	development
	Company	today	experience	world
	Energy	high	gas	africa
	Energy	world	large	alternative
	Energy	record	scale	developing
	Supply	level	transmission	access
	Capability	average	concession	modern
	Company	percent	concession	access
	Energy	real	long	developing
	Resources	world	term	alternative
	Plan	long	pay	dealing
	Resources	term	purchase	principle
	Generation	high	transmission	green
	Energy	percent	demand	develop
	Capacity	plus	focus	alternative
	Operation	low	high	subsidiaries
	Grid	uptime	development	diode
	Grid	corporation	leadership	lighting
	Power	segment	position	highway
	development	segment	practice	lighting
	ownership	segment	type	subsidiary

	distribution	procurement	infrastructure	research
	services	segment	cogeneration	alternative
	services	third	thermal	intellectual
	include	party	subsidiary	specific
	funding	ground	sustainability	develop
	development	perform	biomass	access
	management	financing	traditional	local
	services	efficiency	thermal	improve
	cover	efficiency	gas	clean
	development	efficiency	hybrid	local
	ownership	fast	type	clean
	public	minimize	cogeneration	bulb
	stock	minimize	risk	four
	exchange	perform	component	clean
	subsidiary	two	panorama	five
	owner	rooftop	plant	local
	operator	two	finding	caribbean
	columbia	ten	three	
	history	deliver	turnkey	
	services	megawatt	electric	
	control	maximize	three	
	use	desalinization	turnkey	
	planning		take	
	meet		pay	
	utility		electric	
	high		desalinization	
	remote		generate	
	program		tariff	
	assist		generate	
	remote		putting	
	financing		improve	
	funding			
	access			
	capital			
	markets			
	term			
	formation			
	capital			
	major			
	utility			
	selection			
	near			
	columbia			
	connection			

	north			
	michigan			
	line			
	end			
	interest			
	sigma			
	sigma			
	trading			
	sigma			
	developer			
	sigma			
	competency			
	advantage			
	technical			
	assistance			
	develop			
	diesil			
	thermal			
	origin			
	initiative			
	community			
	identify			
	technical			
	expertise			
	stability			
	technical			
	expertise			
	growing			
	trend			
	future			
	using			
	mix			
	independent			
	develop			
	vancouver			
	two			
	electric			
	trading			
	toronto			
	reputation			
	federal			
	financing			
	access			
	rupert			

	british			
	three			
	international			
	international			
	prince			
	river			

14. The extracted instances from queries when WUP >= 0.5

QI 1	QI 2	QI 3	QI 4	QI 5	QI 6	QI 7
company	generation	wind	wind	engineering	wind	production
power	current	energy	energy	buildings	energy	wind
energy	energy	solarize	energy	energy	energy	energy
power	facilities	countries	wave	household	project	power
wind	range	generation	generation	home	china	power
power	energy	sun	wave	power	power	supply
design	power	power	wind	energy	energy	require
operations	capabilities	buinesses	type	resources	company	average
years	high	supply	fuel	wind	range	amount
require	high	year	speed	water	value	process
high	increase	company	type	sun	cost	type
amount	area	require	fourth	high	type	seventh
area	second	third	query	specification	average	query
type	limit	query	need	nature	consumption	liquid
matter	limit	impact	user	fifth	area	user
average	float	middle	generate	query	sixth	factory
type	query	east			query	
experiment	specific	five			exceed	
kind	given	car				
affect	working	green				
consume	down	positive				
accept	factory					
generate	manager					
exceed						
three						

QI 8	QI 9	QI 10	QI 11	QI 12	QI 13	QI 14	QI 15
engineering	engineering	company	energy	current	energy	businesses	company
wind	energy	water	water	wind	wind	company	energy
energy	wind	resources	power	energy	generation	wind	power

field	company	energy	energy	energy	energy	energy	wind
energy	might	air	company	company	activities	energy	energy
power	offer	current	require	water	businesses	power	energy
value	company	power	type	power	company	energy	company
business	require	energy	average	energy	type	wind	offer
land	type	production	consumption	buildings	area	quality	buildings
eighth	plants	company	float	type	support	company	type
query	services	average	area	average	type	type	average
experiment	major	type	support	consumption	thirteenth	type	consumption
total	nineth	area	type	type	query	close	require
experiment	query	require	eleventh	twelveth	user	area	area
consume		support	query	query		type	distance
far		type	query	query		fourteeenth	type
		tenth	user			query	fifteenth
		query				suggest	query
		consideration				user	query
						query	user
						farm	

QI 16	QI 17	QI 18	QI 19	QI 20	QI 21	QI 22	QI 23
company	company	wind	energy	china	company	wind	water
design	investment	power	water	power	wind	energy	energy
wind	wind	energy	energy	water	energy	air	energy
generation	company	products	power	energy	energy	current	facilities
resources	technologies	provide	technologies	energy	production	energy	water
energy	products	wind	products	supply	power	power	energy
power	services	energy	businesses	capacity	company	energy	technologies
businesses	support	countries	type	water	technologies	facilities	power
average	cover	power	average	design	place	company	power
consumption	area	energy	limit	region	company	support	type
sixteenth	type	company	set	type	type	type	set
query	propose	energy	type	plus	amount	consumption	electricity
megawatt	seventeenth	manufacturer	nineteenth	area	support	type	use
	query	type	query	support	equipment	first	include
	superior	average	cross	type	near	experiment	plants
		type	set	twentieth	first	higher	plants
		eighteenth	time	query	working	cosideration	plants
		query	user	large	twenty	twenty	average
		experiment	consume	scale			consumption
				time			type
				query			flash

				user			suggest
							query
							dry
							steam
							steam
							binary
							cycle
							twenty
							generate

QI 24	QI 25	QI 26	QI 27	QI 28	QI 29	QI 30
company	company	energy	energy	wind	power	company
wind	wind	supply	resources	energy	company	power
power	water	energy	wind	generation	power	resources
wind	power	power	power	wind	energy	power
energy	energy	wind	production	china	company	energy
energy	power	resources	solarize	company	station	energy
supply	energy	wind	range	type	infrastructure	technologies
company	group	design	heat	require	transmission	energy
years	type	wave	electricity	fuel	average	range
type	estimate	company	footprint	vehicle	consumption	development
area	type	type	environment	type	area	alternative
support	area	area	electricity	eighth	support	increase
type	support	support	seventh	query	type	type
fourth	fifth	type	query	apply	nineth	thirtieth
query	query	sixth	case	user	query	query
query	user	query	user	twenty	electric	intellectual
Commit	twenty	user	reduce		query	stable
Twenty	consume	factory	twenty		user	consume
		twenty			twenty	

15. Semantic Matching between instances when WUP >= 0.5

	P1	Human	**Error**	P2	Human	**Error**	P3	Human	**Error**	P4	Human	**Error**
Q 1	32%	70%	**38%**	36%	60%	**24%**	12%	55%	**43%**	24%	65%	**41%**
Q 2	22%	60%	**38%**	30%	70%	**40%**	22%	50%	**28%**	17%	55%	**38%**
Q 3	24%	50%	**26%**	29%	65%	**36%**	9%	40%	**31%**	29%	40%	**11%**
Q 4	19%	55%	**36%**	31%	60%	**29%**	19%	60%	**41%**	25%	50%	**25%**
Q 5	18%	40%	**22%**	35%	55%	**20%**	29%	50%	**21%**	35%	65%	**30%**

Q 6	18%	50%	**32%**	41%	50%	**9%**	24%	55%	**31%**	53%	65%	**12%**
Q 7	25%	45%	**20%**	38%	75%	**37%**	6%	35%	**29%**	38%	75%	**37%**
Q 8	29%	40%	**11%**	43%	60%	**17%**	21%	50%	**29%**	57%	70%	**13%**
Q 9	29%	50%	**21%**	29%	65%	**36%**	29%	65%	**36%**	57%	70%	**13%**
Q 10	26%	65%	**39%**	37%	60%	**23%**	32%	50%	**18%**	42%	55%	**13%**
Q 11	11%	60%	**49%**	24%	75%	**51%**	24%	70%	**46%**	35%	65%	**30%**
Q 12	25%	65%	**40%**	38%	60%	**22%**	44%	70%	**26%**	38%	55%	**17%**
Q 13	21%	60%	**39%**	43%	80%	**37%**	21%	55%	**34%**	36%	65%	**29%**
Q 14	29%	60%	**31%**	33%	60%	**27%**	24%	65%	**41%**	43%	60%	**17%**
Q 15	20%	55%	**35%**	35%	85%	**50%**	30%	60%	**30%**	45%	70%	**25%**
Q 16	41%	60%	**19%**	50%	70%	**20%**	25%	55%	**30%**	41%	60%	**19%**
Q 17	25%	50%	**25%**	25%	65%	**40%**	13%	45%	**32%**	50%	60%	**10%**
Q 18	26%	60%	**34%**	47%	70%	**23%**	32%	65%	**33%**	53%	70%	**17%**
Q 19	11%	50%	**39%**	21%	60%	**39%**	26%	55%	**29%**	21%	50%	**29%**
Q 20	14%	50%	**36%**	18%	55%	**37%**	27%	50%	**23%**	36%	60%	**24%**
Q 21	39%	75%	**36%**	50%	85%	**35%**	17%	40%	**23%**	50%	60%	**10%**
Q 22	22%	60%	**38%**	39%	70%	**31%**	28%	65%	**37%**	39%	55%	**16%**
Q 23	16%	50%	**34%**	19%	75%	**56%**	16%	65%	**49%**	19%	65%	**46%**
Q 24	33%	60%	**27%**	39%	65%	**26%**	22%	60%	**38%**	56%	75%	**19%**
Q 25	22%	65%	**43%**	33%	75%	**42%**	22%	65%	**43%**	44%	70%	**26%**
Q 26	26%	60%	**34%**	32%	65%	**33%**	21%	55%	**34%**	53%	65%	**12%**
Q 27	28%	55%	**27%**	28%	60%	**32%**	17%	35%	**18%**	22%	65%	**43%**
Q 28	25%	60%	**35%**	31%	75%	**44%**	13%	50%	**37%**	31%	70%	**39%**
Q 29	21%	70%	**49%**	26%	85%	**59%**	16%	75%	**59%**	37%	80%	**43%**
Q 30	33%	50%	**17%**	50%	70%	**20%**	22%	60%	**38%**	50%	65%	**15%**

P5	Human	**Error**	P6	Human	**Error**	P7	Human	**Error**	P8	Human	**Error**	P9	Human	**Error**
32%	55%	**23%**	24%	50%	**26%**	40%	65%	**25%**	32%	70%	**38%**	32%	60%	**28%**
26%	50%	**24%**	39%	75%	**36%**	26%	40%	**14%**	26%	50%	**24%**	26%	50%	**24%**
29%	55%	**26%**	29%	60%	**31%**	33%	60%	**27%**	29%	60%	**31%**	29%	50%	**21%**
25%	55%	**30%**	25%	50%	**25%**	13%	20%	**7%**	38%	90%	**52%**	19%	40%	**21%**
41%	65%	**24%**	41%	45%	**4%**	41%	50%	**9%**	47%	65%	**18%**	29%	55%	**26%**
47%	65%	**18%**	41%	60%	**19%**	47%	65%	**18%**	35%	55%	**20%**	35%	60%	**25%**
19%	55%	**36%**	31%	65%	**34%**	31%	55%	**24%**	31%	70%	**39%**	19%	45%	**26%**
50%	75%	**25%**	29%	40%	**11%**	36%	60%	**24%**	43%	60%	**17%**	29%	60%	**31%**
36%	65%	**29%**	43%	60%	**17%**	36%	55%	**19%**	36%	65%	**29%**	29%	55%	**26%**
32%	70%	**38%**	37%	75%	**38%**	37%	50%	**13%**	37%	60%	**23%**	32%	50%	**18%**
29%	80%	**51%**	29%	85%	**56%**	29%	75%	**46%**	29%	90%	**61%**	29%	75%	**46%**
38%	75%	**37%**	38%	60%	**22%**	44%	60%	**16%**	44%	70%	**26%**	31%	60%	**29%**
29%	70%	**41%**	36%	65%	**29%**	29%	50%	**21%**	50%	90%	**40%**	36%	55%	**19%**
38%	75%	**37%**	33%	65%	**32%**	33%	60%	**27%**	33%	75%	**42%**	38%	85%	**47%**

30%	75%	**45%**	30%	60%	**30%**	40%	70%	**30%**	30%	70%	**40%**	35%	70%	**35%**
33%	70%	**37%**	58%	75%	**17%**	25%	40%	**15%**	58%	80%	**22%**	25%	50%	**25%**
50%	80%	**30%**	31%	50%	**19%**	31%	55%	**24%**	25%	60%	**35%**	44%	75%	**31%**
42%	75%	**33%**	37%	60%	**23%**	53%	70%	**17%**	47%	80%	**33%**	53%	70%	**17%**
42%	85%	**43%**	26%	65%	**39%**	32%	60%	**28%**	26%	65%	**39%**	32%	80%	**48%**
40%	80%	**40%**	36%	80%	**44%**	14%	40%	**26%**	18%	50%	**32%**	27%	40%	**13%**
39%	65%	**26%**	39%	55%	**16%**	44%	75%	**31%**	44%	70%	**26%**	44%	60%	**16%**
28%	65%	**37%**	33%	65%	**32%**	28%	60%	**32%**	28%	60%	**32%**	33%	45%	**12%**
32%	75%	**43%**	32%	75%	**43%**	23%	50%	**27%**	16%	50%	**34%**	26%	55%	**29%**
33%	60%	**27%**	33%	60%	**27%**	33%	65%	**32%**	44%	80%	**36%**	39%	60%	**21%**
33%	75%	**42%**	33%	80%	**47%**	28%	75%	**47%**	39%	85%	**46%**	33%	65%	**32%**
26%	60%	**34%**	32%	55%	**23%**	21%	50%	**29%**	37%	65%	**28%**	33%	30%	**3%**
33%	50%	**17%**	22%	50%	**28%**	33%	60%	**27%**	44%	75%	**31%**	17%	35%	**18%**
25%	65%	**40%**	19%	40%	**21%**	13%	30%	**17%**	31%	75%	**44%**	13%	30%	**17%**
26%	85%	**59%**	26%	80%	**54%**	32%	85%	**53%**	37%	90%	**53%**	32%	80%	**48%**
50%	85%	**35%**	50%	85%	**35%**	33%	60%	**27%**	50%	80%	**30%**	63%	75%	**12%**

16. The extracted instances from providers with KAON

P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**
power	power	water	offer	power	engineering	power	engineering	offer
capacity	value	design	power	capacity	water	project	power	power
design	order	engineering	value	water	project	range	water	range
engineering	lead		capacity	value	power	engineering		home
			engineering	project	capacity			
			supply	engineering	design			
				design	range			
				field	supply			

17. The extracted instances from queries with KAON

QI 1	QI 2	QI 3	QI 4	QI 5	QI 6	QI 7	QI 8	QI 9	QI 10
power	range	power		engineering	project	power	engineering	engineering	water
power	power	supply		home	power	power	field	offer	power
power				power	range	supply	power		
design				water	value		value		

QI 11	QI 12	QI 13	QI 14	QI 15	QI 16	QI 17	QI 18	QI 19	QI 20
water	water		power	power	design		power	water	power

power	power			offer	power		power	power	water
									supply
									capacity
									water
									design

QI 21	QI 22	QI 23	QI 24	QI 25	QI 26	QI 27	QI 28	QI 29	QI 30
power	power	water	power	water	supply	power		power	power
place		water	supply	power	power	range		power	power
		power		power	design				range
		power							

	P1	Human	**Error**	P2	Human	**Error**	P3	Human	**Error**	P4	Human	**Error**
Q1	100%	70%	**30%**	75%	60%	**15%**	25%	55%	**30%**	75%	65%	**10%**
Q2	50%	60%	**10%**	50%	70%	**20%**	0%	50%	**50%**	50%	55%	**5%**
Q3	50%	50%	**0%**	50%	65%	**15%**	0%	40%	**40%**	100%	40%	**60%**
Q4	0%	55%	**55%**	0%	60%	**60%**	0%	60%	**60%**	0%	50%	**50%**
Q5	50%	40%	**10%**	25%	55%	**30%**	50%	50%	**0%**	50%	65%	**15%**
Q6	25%	50%	**25%**	50%	50%	**0%**	0%	55%	**55%**	50%	65%	**15%**
Q7	67%	45%	**22%**	67%	75%	**8%**	0%	35%	**35%**	100%	75%	**25%**
Q8	50%	40%	**10%**	50%	60%	**10%**	25%	50%	**25%**	75%	70%	**5%**
Q9	50%	50%	**0%**	0%	65%	**65%**	50%	65%	**15%**	100%	70%	**30%**
Q10	50%	65%	**15%**	50%	60%	**10%**	50%	50%	**0%**	50%	55%	**5%**
Q11	50%	60%	**10%**	50%	75%	**25%**	50%	70%	**20%**	50%	65%	**15%**
Q12	50%	65%	**15%**	50%	60%	**10%**	50%	70%	**20%**	50%	55%	**5%**
Q13	0%	60%	**60%**	0%	80%	**80%**	0%	55%	**55%**	0%	65%	**65%**
Q14	100%	60%	**40%**	100%	60%	**40%**	0%	65%	**65%**	100%	60%	**40%**
Q15	50%	55%	**5%**	50%	85%	**35%**	0%	60%	**60%**	100%	70%	**30%**
Q16	100%	60%	**40%**	50%	70%	**20%**	50%	55%	**5%**	50%	60%	**10%**
Q17	0%	50%	**50%**	0%	65%	**65%**	0%	45%	**45%**	0%	60%	**60%**
Q18	100%	60%	**40%**	100%	70%	**30%**	0%	65%	**65%**	100%	70%	**30%**
Q19	50%	50%	**0%**	50%	60%	**10%**	50%	55%	**5%**	50%	50%	**0%**
Q20	50%	50%	**0%**	17%	55%	**38%**	50%	50%	**0%**	33%	60%	**27%**
Q21	50%	75%	**25%**	50%	85%	**35%**	0%	40%	**40%**	50%	60%	**10%**
Q22	100%	60%	**40%**	100%	70%	**30%**	0%	65%	**65%**	100%	55%	**45%**
Q23	50%	50%	**0%**	50%	75%	**25%**	50%	65%	**15%**	50%	65%	**15%**
Q24	67%	60%	**7%**	67%	65%	**2%**	33%	60%	**27%**	67%	75%	**8%**
Q25	67%	65%	**2%**	67%	75%	**8%**	33%	65%	**32%**	67%	70%	**3%**

Q26	67%	60%	**7%**	33%	65%	**32%**	33%	55%	**22%**	67%	65%	**2%**
Q27	50%	55%	**5%**	50%	60%	**10%**	0%	35%	**35%**	50%	65%	**15%**
Q28	0%	60%	**60%**	0%	75%	**75%**	0%	50%	**50%**	0%	70%	**70%**
Q29	100%	70%	**30%**	100%	85%	**15%**	0%	75%	**75%**	100%	80%	**20%**
Q30	67%	50%	**17%**	67%	70%	**3%**	0%	60%	**60%**	67%	65%	**2%**

P5	Human	**Error**	P6	Human	**Error**	P7	Human	**Error**
100%	55%	**45%**	100%	50%	**50%**	75%	65%	**10%**
50%	50%	**0%**	100%	75%	**25%**	100%	40%	**60%**
50%	55%	**5%**	100%	60%	**40%**	50%	60%	**10%**
0%	55%	**55%**	0%	50%	**50%**	0%	20%	**20%**
75%	65%	**10%**	75%	45%	**30%**	50%	50%	**0%**
75%	65%	**10%**	75%	60%	**15%**	75%	65%	**10%**
67%	55%	**12%**	100%	65%	**35%**	67%	55%	**12%**
100%	75%	**25%**	50%	40%	**10%**	50%	60%	**10%**
50%	65%	**15%**	50%	60%	**10%**	50%	55%	**5%**
100%	70%	**30%**	100%	75%	**25%**	50%	50%	**0%**
100%	80%	**20%**	100%	85%	**15%**	50%	75%	**25%**
100%	75%	**25%**	100%	60%	**40%**	50%	60%	**10%**
0%	70%	**70%**	0%	65%	**65%**	0%	50%	**50%**
100%	75%	**25%**	100%	65%	**35%**	100%	60%	**40%**
50%	75%	**25%**	50%	60%	**10%**	50%	70%	**20%**
100%	70%	**30%**	100%	75%	**25%**	50%	40%	**10%**
0%	80%	**80%**	0%	50%	**50%**	0%	55%	**55%**
100%	75%	**25%**	100%	60%	**40%**	100%	70%	**30%**
100%	85%	**15%**	100%	65%	**35%**	50%	60%	**10%**
83%	80%	**3%**	100%	80%	**20%**	17%	40%	**23%**
50%	65%	**15%**	50%	55%	**5%**	50%	75%	**25%**
100%	65%	**35%**	100%	65%	**35%**	100%	60%	**40%**
100%	75%	**25%**	100%	75%	**25%**	50%	50%	**0%**
100%	60%	**40%**	100%	60%	**40%**	67%	65%	**2%**
100%	75%	**25%**	100%	80%	**20%**	67%	75%	**8%**
67%	60%	**7%**	67%	55%	**12%**	33%	50%	**17%**
50%	50%	**0%**	100%	50%	**50%**	100%	60%	**40%**
0%	65%	**65%**	0%	40%	**40%**	0%	30%	**30%**
100%	85%	**15%**	100%	80%	**20%**	100%	85%	**15%**
67%	85%	**18%**	100%	85%	**15%**	100%	60%	**40%**

P8	Human	**Error**	P9	Human	**Error**
75%	70%	**5%**	75%	60%	**15%**
50%	50%	**0%**	100%	50%	**50%**
50%	60%	**10%**	50%	50%	**0%**
0%	90%	**90%**	0%	40%	**40%**
75%	65%	**10%**	50%	55%	**5%**
25%	55%	**30%**	50%	60%	**10%**
67%	70%	**3%**	67%	45%	**22%**
50%	60%	**10%**	25%	60%	**35%**
50%	65%	**15%**	50%	55%	**5%**
100%	60%	**40%**	50%	50%	**0%**
100%	90%	**10%**	50%	75%	**25%**
100%	70%	**30%**	50%	60%	**10%**
0%	90%	**90%**	0%	55%	**55%**
100%	75%	**25%**	100%	85%	**15%**
50%	70%	**20%**	100%	70%	**30%**
50%	80%	**30%**	50%	50%	**0%**
0%	60%	**60%**	0%	75%	**75%**
100%	80%	**20%**	100%	70%	**30%**
100%	65%	**35%**	50%	80%	**30%**
50%	50%	**0%**	17%	40%	**23%**
50%	70%	**20%**	50%	60%	**10%**
100%	60%	**40%**	100%	45%	**55%**
100%	50%	**50%**	50%	55%	**5%**
100%	80%	**20%**	67%	60%	**7%**
100%	85%	**15%**	67%	65%	**2%**
33%	65%	**32%**	33%	30%	**3%**
50%	75%	**25%**	100%	35%	**65%**
0%	75%	**75%**	0%	30%	**30%**
100%	90%	**10%**	100%	80%	**20%**
67%	80%	**13%**	100%	75%	**25%**

18. Semantic Matching after Frequency Normalization with WUP >= 0.9

	P1	Human	Error	P2	Human	Error	P3	Human	Error
Q1	**89%**	70%	**19%**	**68%**	60%	**8%**	**34%**	55%	**21%**
Q2	**43%**	60%	**17%**	**72%**	70%	**2%**	**38%**	50%	**12%**
Q3	**49%**	50%	**1%**	**61%**	65%	**4%**	**25%**	40%	**15%**
Q4	**56%**	55%	**1%**	**94%**	60%	**34%**	**56%**	60%	**4%**

Q5	**43%**	40%	**3%**	**43%**	55%	**12%**	**48%**	50%	**2%**
Q6	**30%**	50%	**20%**	**70%**	50%	**20%**	**40%**	55%	**15%**
Q7	**73%**	45%	**28%**	**91%**	75%	**16%**	**18%**	35%	**17%**
Q8	**42%**	40%	**2%**	**70%**	60%	**10%**	**42%**	50%	**8%**
Q9	**65%**	50%	**15%**	**65%**	65%	**0%**	**65%**	65%	**0%**
Q10	**51%**	65%	**14%**	**66%**	60%	**6%**	**62%**	50%	**12%**
Q11	**40%**	60%	**20%**	**80%**	75%	**5%**	**80%**	70%	**10%**
Q12	**45%**	65%	**20%**	**74%**	60%	**14%**	**75%**	70%	**5%**
Q13	**44%**	60%	**16%**	**85%**	80%	**5%**	**44%**	55%	**11%**
Q14	**60%**	60%	**0%**	**80%**	60%	**20%**	**50%**	65%	**15%**
Q15	**47%**	55%	**8%**	**82%**	85%	**3%**	**65%**	60%	**5%**
Q16	**63%**	60%	**3%**	**75%**	70%	**5%**	**25%**	55%	**30%**
Q17	**77%**	50%	**27%**	**63%**	65%	**2%**	**40%**	45%	**5%**
Q18	**45%**	60%	**15%**	**81%**	70%	**11%**	**48%**	65%	**17%**
Q19	**27%**	50%	**23%**	**44%**	60%	**16%**	**44%**	55%	**11%**
Q20	**36%**	50%	**14%**	**36%**	55%	**19%**	**60%**	50%	**10%**
Q21	**76%**	75%	**1%**	**89%**	85%	**4%**	**56%**	40%	**16%**
Q22	**45%**	60%	**15%**	**86%**	70%	**16%**	**57%**	65%	**8%**
Q23	**32%**	50%	**18%**	**68%**	75%	**7%**	**57%**	65%	**8%**
Q24	**70%**	60%	**10%**	**84%**	65%	**19%**	**48%**	60%	**12%**
Q25	**57%**	65%	**8%**	**86%**	75%	**11%**	**57%**	65%	**8%**
Q26	**58%**	60%	**2%**	**92%**	65%	**27%**	**46%**	55%	**9%**
Q27	**47%**	55%	**8%**	**79%**	60%	**19%**	**16%**	35%	**19%**
Q28	**69%**	60%	**9%**	**86%**	75%	**11%**	**51%**	50%	**1%**
Q29	**80%**	70%	**10%**	**100%**	85%	**15%**	**60%**	75%	**15%**
Q30	**43%**	50%	**7%**	**79%**	70%	**9%**	**45%**	60%	**15%**

P4	Human	Error	P5	Huamn	Error	P6	Human	Error
57%	65%	**8%**	**66%**	55%	**11%**	**68%**	50%	**18%**
43%	55%	**12%**	**43%**	50%	**7%**	**86%**	75%	**11%**
43%	40%	**3%**	**51%**	55%	**4%**	**69%**	60%	**9%**
75%	50%	**25%**	**37%**	55%	**18%**	**56%**	50%	**6%**
64%	65%	**1%**	**54%**	65%	**11%**	**64%**	45%	**19%**
78%	65%	**13%**	**78%**	65%	**13%**	**70%**	60%	**10%**
100%	75%	**25%**	**55%**	55%	**0%**	**64%**	65%	**1%**
98%	70%	**28%**	**72%**	75%	**3%**	**56%**	40%	**16%**
97%	70%	**27%**	**65%**	65%	**0%**	**65%**	60%	**5%**
76%	55%	**21%**	**66%**	70%	**4%**	**87%**	75%	**12%**
80%	65%	**15%**	**100%**	80%	**20%**	**100%**	85%	**15%**
74%	55%	**19%**	**74%**	75%	**1%**	**74%**	60%	**14%**
59%	65%	**6%**	**59%**	70%	**11%**	**73%**	65%	**8%**
90%	60%	**30%**	**80%**	75%	**5%**	**70%**	65%	**5%**

94%	70%	**24%**	**71%**	75%	**4%**	**71%**	60%	**11%**
63%	60%	**3%**	**50%**	70%	**20%**	**75%**	75%	**0%**
80%	60%	**20%**	**80%**	80%	**0%**	**40%**	50%	**10%**
90%	70%	**20%**	**72%**	75%	**3%**	**63%**	60%	**3%**
59%	50%	**9%**	**85%**	85%	**0%**	**73%**	65%	**8%**
64%	60%	**4%**	**70%**	80%	**10%**	**90%**	80%	**10%**
89%	60%	**29%**	**76%**	65%	**11%**	**67%**	55%	**12%**
74%	55%	**19%**	**61%**	65%	**4%**	**74%**	65%	**9%**
57%	65%	**8%**	**89%**	75%	**14%**	**68%**	75%	**7%**
90%	75%	**15%**	**60%**	60%	**0%**	**66%**	60%	**6%**
86%	70%	**16%**	**86%**	75%	**11%**	**86%**	80%	**6%**
75%	65%	**10%**	**58%**	60%	**2%**	**69%**	55%	**14%**
79%	65%	**14%**	**32%**	50%	**18%**	**47%**	50%	**3%**
83%	70%	**13%**	**49%**	65%	**16%**	**51%**	40%	**11%**
100%	80%	**20%**	**100%**	85%	**15%**	**100%**	80%	**20%**
79%	65%	**14%**	**77%**	85%	**8%**	**91%**	85%	**6%**

P7	Human	Error	P8	Human	Error	P9	Human	Error
77%	65%	**12%**	**66%**	70%	**4%**	**66%**	60%	**6%**
57%	40%	**17%**	**57%**	50%	**7%**	**57%**	50%	**7%**
63%	60%	**3%**	**74%**	60%	**14%**	**59%**	50%	**9%**
37%	20%	**17%**	**94%**	90%	**4%**	**37%**	40%	**3%**
52%	50%	**2%**	**54%**	65%	**11%**	**48%**	55%	**7%**
70%	65%	**5%**	**60%**	55%	**5%**	**60%**	60%	**0%**
73%	55%	**18%**	**73%**	70%	**3%**	**55%**	45%	**10%**
70%	60%	**10%**	**84%**	60%	**24%**	**56%**	60%	**4%**
65%	55%	**10%**	**81%**	65%	**16%**	**65%**	55%	**10%**
66%	50%	**16%**	**87%**	60%	**27%**	**55%**	50%	**5%**
100%	75%	**25%**	**100%**	90%	**10%**	**80%**	75%	**5%**
80%	60%	**20%**	**86%**	70%	**16%**	**61%**	60%	**1%**
59%	50%	**9%**	**100%**	90%	**10%**	**59%**	55%	**4%**
70%	60%	**10%**	**90%**	75%	**15%**	**80%**	85%	**5%**
76%	70%	**6%**	**71%**	70%	**1%**	**82%**	70%	**12%**
38%	40%	**3%**	**75%**	80%	**5%**	**38%**	50%	**13%**
60%	55%	**5%**	**60%**	60%	**0%**	**80%**	75%	**5%**
79%	70%	**9%**	**81%**	80%	**1%**	**82%**	70%	**12%**
73%	60%	**13%**	**59%**	65%	**6%**	**85%**	80%	**5%**
36%	40%	**4%**	**60%**	50%	**10%**	**36%**	40%	**4%**
78%	75%	**3%**	**89%**	70%	**19%**	**76%**	60%	**16%**
61%	60%	**1%**	**74%**	60%	**14%**	**61%**	45%	**16%**
57%	50%	**7%**	**79%**	50%	**29%**	**66%**	55%	**11%**
70%	65%	**5%**	**82%**	80%	**2%**	**70%**	60%	**10%**

71%	75%	**4%**	**100%**	85%	**15%**	**71%**	65%	**6%**
46%	50%	**4%**	**81%**	65%	**16%**	**46%**	30%	**16%**
63%	60%	**3%**	**79%**	75%	**4%**	**47%**	35%	**12%**
34%	30%	**4%**	**86%**	75%	**11%**	**34%**	30%	**4%**
100%	85%	**15%**	**100%**	90%	**10%**	**100%**	80%	**20%**
79%	60%	**19%**	**79%**	80%	**1%**	**89%**	75%	**14%**

19. Semantic Matching after Frequency Norm. with WUP >= 0.8

	P1	Human	Error	P2	Human	Error	P3	Human	Error
Q1	65%	70%	5%	51%	60%	9%	24%	55%	31%
Q2	40%	60%	20%	49%	70%	21%	26%	50%	24%
Q3	40%	50%	10%	50%	65%	15%	20%	40%	20%
Q4	43%	55%	12%	72%	60%	12%	43%	60%	17%
Q5	26%	40%	14%	37%	55%	18%	43%	50%	7%
Q6	26%	50%	24%	60%	50%	10%	34%	55%	21%
Q7	52%	45%	7%	65%	75%	10%	13%	35%	22%
Q8	39%	40%	1%	78%	60%	18%	39%	50%	11%
Q9	50%	50%	0%	50%	65%	15%	50%	65%	15%
Q10	38%	65%	27%	46%	60%	14%	46%	50%	4%
Q11	23%	60%	37%	46%	75%	29%	46%	70%	24%
Q12	39%	65%	26%	48%	60%	12%	58%	70%	12%
Q13	32%	60%	28%	64%	80%	16%	32%	55%	23%
Q14	47%	60%	13%	63%	60%	3%	39%	65%	26%
Q15	33%	55%	22%	58%	85%	27%	50%	60%	10%
Q16	58%	60%	2%	69%	70%	1%	23%	55%	32%
Q17	44%	50%	6%	44%	65%	21%	22%	45%	23%
Q18	38%	60%	22%	69%	70%	1%	44%	65%	21%
Q19	23%	50%	27%	35%	60%	25%	35%	55%	20%
Q20	25%	50%	25%	25%	55%	30%	42%	50%	8%
Q21	62%	75%	13%	74%	85%	11%	44%	40%	4%
Q22	38%	60%	22%	66%	70%	4%	47%	65%	18%
Q23	30%	50%	20%	45%	75%	30%	38%	65%	27%
Q24	53%	60%	7%	53%	65%	12%	35%	60%	25%
Q25	40%	65%	25%	60%	75%	15%	40%	65%	25%
Q26	43%	60%	17%	69%	65%	4%	34%	55%	21%
Q27	50%	55%	5%	50%	60%	10%	10%	35%	25%

Q28	52%	60%	8%	52%	75%	23%	26%	50%	24%
Q29	46%	70%	24%	58%	85%	27%	35%	75%	40%
Q30	51%	50%	1%	69%	70%	1%	34%	60%	26%

P4	Human	Error	P5	Huamn	Error	P6	Human	Error
57%	65%	8%	51%	55%	4%	51%	50%	1%
34%	55%	21%	40%	50%	10%	57%	75%	18%
40%	40%	0%	43%	55%	12%	63%	60%	3%
57%	50%	7%	33%	55%	22%	43%	50%	7%
51%	65%	14%	46%	65%	19%	54%	45%	9%
77%	65%	12%	63%	65%	2%	54%	60%	6%
78%	75%	3%	39%	55%	16%	52%	65%	13%
91%	70%	21%	70%	75%	5%	52%	40%	12%
75%	70%	5%	50%	65%	15%	50%	60%	10%
67%	55%	12%	46%	70%	24%	62%	75%	13%
66%	65%	1%	58%	80%	22%	58%	85%	27%
48%	55%	7%	48%	75%	27%	48%	60%	12%
61%	65%	4%	43%	70%	27%	54%	65%	11%
79%	60%	19%	63%	75%	12%	55%	65%	10%
75%	70%	5%	50%	75%	25%	50%	60%	10%
58%	60%	2%	46%	70%	24%	69%	75%	6%
63%	60%	3%	56%	80%	24%	22%	50%	28%
77%	70%	7%	62%	75%	13%	54%	60%	6%
46%	50%	4%	69%	85%	16%	58%	65%	7%
64%	60%	4%	50%	80%	30%	67%	80%	13%
77%	60%	17%	62%	65%	3%	53%	55%	2%
63%	55%	8%	47%	65%	18%	56%	65%	9%
40%	65%	25%	68%	75%	7%	63%	75%	12%
85%	75%	10%	44%	60%	16%	53%	60%	7%
77%	70%	7%	60%	75%	15%	60%	80%	20%
74%	65%	9%	43%	60%	17%	51%	55%	4%
50%	65%	15%	40%	50%	10%	30%	50%	20%
65%	70%	5%	44%	65%	21%	39%	40%	1%
77%	80%	3%	58%	85%	27%	58%	80%	22%
60%	65%	5%	69%	85%	16%	69%	85%	16%

P7	Human	Error	P8	Human	Error	P9	Human	Error
62%	65%	3%	51%	70%	19%	49%	60%	11%
40%	40%	0%	40%	50%	10%	43%	50%	7%
57%	60%	3%	60%	60%	0%	50%	50%	0%
29%	20%	9%	72%	90%	18%	29%	40%	11%
51%	50%	1%	63%	65%	2%	43%	55%	12%
57%	65%	8%	52%	55%	3%	52%	60%	8%
57%	55%	2%	65%	70%	5%	39%	45%	6%
65%	60%	5%	78%	60%	18%	52%	60%	8%
50%	55%	5%	63%	65%	2%	50%	55%	5%
49%	50%	1%	62%	60%	2%	44%	50%	6%
62%	75%	13%	58%	90%	32%	42%	75%	33%
61%	60%	1%	58%	70%	12%	39%	60%	21%
43%	50%	7%	75%	90%	15%	50%	55%	5%
55%	60%	5%	71%	75%	4%	63%	85%	22%
61%	70%	9%	58%	70%	12%	58%	70%	12%
38%	40%	2%	69%	80%	11%	35%	50%	15%
33%	55%	22%	33%	60%	27%	63%	75%	12%
72%	70%	2%	69%	80%	11%	74%	70%	4%
62%	60%	2%	46%	65%	19%	69%	80%	11%
25%	40%	15%	42%	50%	8%	31%	40%	9%
62%	75%	13%	71%	70%	1%	68%	60%	8%
47%	60%	13%	56%	60%	4%	53%	45%	8%
45%	50%	5%	60%	50%	10%	48%	55%	7%
53%	65%	12%	71%	80%	9%	59%	60%	1%
50%	75%	25%	70%	85%	15%	57%	65%	8%
34%	50%	16%	60%	65%	5%	40%	30%	10%
60%	60%	0%	70%	75%	5%	30%	35%	5%
26%	30%	4%	65%	75%	10%	26%	30%	4%
62%	85%	23%	58%	90%	32%	66%	80%	14%
60%	60%	0%	60%	80%	20%	77%	75%	2%

20. Semantic Matching after Frequency Norm. with WUP >= 0.7

	P1	Human	Error	P2	Human	Error	P3	Human	Error
Q1	60%	70%	10%	48%	60%	12%	23%	55%	32%

Q2	35%	60%	25%	43%	70%	27%	23%	50%	27%
Q3	33%	50%	17%	42%	65%	23%	17%	40%	23%
Q4	36%	55%	19%	60%	60%	0%	36%	60%	24%
Q5	24%	40%	16%	35%	55%	20%	41%	50%	9%
Q6	24%	50%	26%	57%	50%	7%	32%	55%	23%
Q7	45%	45%	0%	56%	75%	19%	11%	35%	24%
Q8	32%	40%	8%	64%	60%	4%	32%	50%	18%
Q9	46%	50%	4%	46%	65%	19%	46%	65%	19%
Q10	36%	65%	29%	43%	60%	17%	43%	50%	7%
Q11	20%	60%	40%	40%	75%	35%	40%	70%	30%
Q12	35%	65%	30%	44%	60%	16%	53%	70%	17%
Q13	29%	60%	31%	58%	80%	22%	29%	55%	26%
Q14	42%	60%	18%	56%	60%	4%	35%	65%	30%
Q15	30%	55%	25%	53%	85%	32%	45%	60%	15%
Q16	54%	60%	6%	64%	70%	6%	21%	55%	34%
Q17	40%	50%	10%	40%	65%	25%	20%	45%	25%
Q18	36%	60%	24%	64%	70%	6%	41%	65%	24%
Q19	19%	50%	31%	28%	60%	32%	28%	55%	27%
Q20	21%	50%	29%	21%	55%	34%	35%	50%	15%
Q21	58%	75%	17%	70%	85%	15%	42%	40%	2%
Q22	33%	60%	27%	58%	70%	12%	42%	65%	23%
Q23	28%	50%	22%	42%	75%	33%	35%	65%	30%
Q24	49%	60%	11%	49%	65%	16%	32%	60%	28%
Q25	36%	65%	29%	55%	75%	20%	36%	65%	29%
Q26	40%	60%	20%	63%	65%	2%	32%	55%	23%
Q27	44%	55%	11%	44%	60%	16%	9%	35%	26%
Q28	45%	60%	15%	45%	75%	30%	22%	50%	28%
Q29	39%	70%	31%	49%	85%	36%	29%	75%	46%
Q30	46%	50%	4%	62%	70%	8%	31%	60%	29%

P4	Human	Error	P5	Huamn	Error	P6	Human	Error
53%	65%	12%	48%	55%	7%	48%	50%	2%
30%	55%	25%	35%	50%	15%	50%	75%	25%
33%	40%	7%	36%	55%	19%	53%	60%	7%
48%	50%	2%	28%	55%	27%	36%	50%	14%
49%	65%	16%	43%	65%	22%	51%	45%	6%
73%	65%	8%	60%	65%	5%	51%	60%	9%
67%	75%	8%	33%	55%	22%	45%	65%	20%
75%	70%	5%	57%	75%	18%	43%	40%	3%
69%	70%	1%	46%	65%	19%	46%	60%	14%
62%	55%	7%	43%	70%	27%	57%	75%	18%
57%	65%	8%	50%	80%	30%	50%	85%	35%

44%	55%	11%	44%	75%	31%	44%	60%	16%
55%	65%	10%	39%	70%	31%	48%	65%	17%
70%	60%	10%	56%	75%	19%	49%	65%	16%
68%	70%	2%	45%	75%	30%	45%	60%	15%
54%	60%	6%	43%	70%	27%	64%	75%	11%
57%	60%	3%	50%	80%	30%	20%	50%	30%
72%	70%	2%	57%	75%	18%	50%	60%	10%
38%	50%	12%	56%	85%	29%	47%	65%	18%
54%	60%	6%	42%	80%	38%	56%	80%	24%
72%	60%	12%	58%	65%	7%	50%	55%	5%
56%	55%	1%	42%	65%	23%	50%	65%	15%
37%	65%	28%	63%	75%	12%	58%	75%	17%
78%	75%	3%	41%	60%	19%	49%	60%	11%
70%	70%	0%	55%	75%	20%	55%	80%	25%
69%	65%	4%	40%	60%	20%	47%	55%	8%
44%	65%	21%	35%	50%	15%	27%	50%	23%
56%	70%	14%	37%	65%	28%	33%	40%	7%
65%	80%	15%	49%	85%	36%	49%	80%	31%
54%	65%	11%	62%	85%	23%	62%	85%	23%

P7	Human	Error	P8	Human	Error	P9	Human	Error
58%	65%	7%	48%	70%	22%	45%	60%	15%
35%	40%	5%	35%	50%	15%	38%	50%	12%
47%	60%	13%	50%	60%	10%	42%	50%	8%
24%	20%	4%	60%	90%	30%	24%	40%	16%
49%	50%	1%	59%	65%	6%	41%	55%	14%
54%	65%	11%	49%	55%	6%	49%	60%	11%
48%	55%	7%	56%	70%	14%	33%	45%	12%
54%	60%	6%	64%	60%	4%	43%	60%	17%
46%	55%	9%	58%	65%	7%	46%	55%	9%
45%	50%	5%	57%	60%	3%	41%	50%	9%
53%	75%	22%	50%	90%	40%	37%	75%	38%
56%	60%	4%	53%	70%	17%	35%	60%	25%
39%	50%	11%	68%	90%	22%	45%	55%	10%
49%	60%	11%	63%	75%	12%	56%	85%	29%
55%	70%	15%	53%	70%	17%	53%	70%	17%
36%	40%	4%	64%	80%	16%	32%	50%	18%
30%	55%	25%	30%	60%	30%	57%	75%	18%
67%	70%	3%	64%	80%	16%	69%	70%	1%
50%	60%	10%	38%	65%	27%	56%	80%	24%
21%	40%	19%	35%	50%	15%	26%	40%	14%
58%	75%	17%	67%	70%	3%	64%	60%	4%

42%	60%	18%	50%	60%	10%	47%	45%	2%
42%	50%	8%	56%	50%	6%	44%	55%	11%
49%	65%	16%	65%	80%	15%	54%	60%	6%
46%	75%	29%	64%	85%	21%	52%	65%	13%
32%	50%	18%	55%	65%	10%	37%	30%	7%
53%	60%	7%	62%	75%	13%	27%	35%	8%
22%	30%	8%	56%	75%	19%	22%	30%	8%
52%	85%	33%	49%	90%	41%	55%	80%	25%
54%	60%	6%	54%	80%	26%	69%	75%	6%

21. Semantic Matching after Frequency Normalization with WUP >= 0.6

	P1	Human	Error	P2	Human	Error	P3	Human	Error
Q1	59%	70%	11%	46%	60%	14%	22%	55%	33%
Q2	33%	60%	27%	41%	70%	29%	21%	50%	29%
Q3	31%	50%	19%	39%	65%	26%	15%	40%	25%
Q4	36%	55%	19%	60%	60%	0%	36%	60%	24%
Q5	24%	40%	16%	35%	55%	20%	41%	50%	9%
Q6	24%	50%	26%	57%	50%	7%	32%	55%	23%
Q7	43%	45%	2%	54%	75%	21%	11%	35%	24%
Q8	32%	40%	8%	64%	60%	4%	32%	50%	18%
Q9	46%	50%	4%	46%	65%	19%	46%	65%	19%
Q10	36%	65%	29%	43%	60%	17%	43%	50%	7%
Q11	20%	60%	40%	40%	75%	35%	40%	70%	30%
Q12	35%	65%	30%	44%	60%	16%	53%	70%	17%
Q13	29%	60%	31%	58%	80%	22%	29%	55%	26%
Q14	42%	60%	18%	56%	60%	4%	35%	65%	30%
Q15	30%	55%	25%	53%	85%	32%	45%	60%	15%
Q16	54%	60%	6%	64%	70%	6%	21%	55%	34%
Q17	40%	50%	10%	40%	65%	25%	20%	45%	25%
Q18	36%	60%	24%	64%	70%	6%	41%	65%	24%
Q19	19%	50%	31%	28%	60%	32%	28%	55%	27%
Q20	21%	50%	29%	21%	55%	34%	35%	50%	15%
Q21	58%	75%	17%	70%	85%	15%	42%	40%	2%
Q22	33%	60%	27%	58%	70%	12%	42%	65%	23%
Q23	28%	50%	22%	42%	75%	33%	35%	65%	30%
Q24	47%	60%	13%	47%	65%	18%	32%	60%	28%
Q25	36%	65%	29%	55%	75%	20%	36%	65%	29%
Q26	39%	60%	21%	62%	65%	3%	31%	55%	24%
Q27	43%	55%	12%	43%	60%	17%	9%	35%	26%

Q28	45%	60%	15%	45%	75%	30%	22%	50%	28%
Q29	39%	70%	31%	49%	85%	36%	29%	75%	46%
Q30	46%	50%	4%	62%	70%	8%	31%	60%	29%

P4	Human	Error	P5	Human	Error	P6	Human	Error
51%	65%	14%	46%	55%	9%	46%	50%	4%
29%	55%	26%	33%	50%	17%	48%	75%	27%
31%	40%	9%	33%	55%	22%	49%	60%	11%
48%	50%	2%	28%	55%	27%	36%	50%	14%
49%	65%	16%	43%	65%	22%	51%	45%	6%
73%	65%	8%	60%	65%	5%	51%	60%	9%
65%	75%	10%	32%	55%	23%	43%	65%	22%
75%	70%	5%	57%	75%	18%	43%	40%	3%
69%	70%	1%	46%	65%	19%	46%	60%	14%
62%	55%	7%	43%	70%	27%	57%	75%	18%
57%	65%	8%	50%	80%	30%	50%	85%	35%
44%	55%	11%	44%	75%	31%	44%	60%	16%
55%	65%	10%	39%	70%	31%	48%	65%	17%
70%	60%	10%	56%	75%	19%	49%	65%	16%
68%	70%	2%	45%	75%	30%	45%	60%	15%
54%	60%	6%	43%	70%	27%	64%	75%	11%
57%	60%	3%	50%	80%	30%	20%	50%	30%
72%	70%	2%	57%	75%	18%	50%	60%	10%
38%	50%	12%	56%	85%	29%	47%	65%	18%
54%	60%	6%	42%	80%	38%	56%	80%	24%
72%	60%	12%	58%	65%	7%	50%	55%	5%
56%	55%	1%	42%	65%	23%	50%	65%	15%
37%	65%	28%	63%	75%	12%	58%	75%	17%
76%	75%	1%	40%	60%	20%	47%	60%	13%
70%	70%	0%	55%	75%	20%	55%	80%	25%
67%	65%	2%	39%	60%	21%	46%	55%	9%
43%	65%	22%	34%	50%	16%	26%	50%	24%
56%	70%	14%	37%	65%	28%	33%	40%	7%
65%	80%	15%	49%	85%	36%	49%	80%	31%
54%	65%	11%	62%	85%	23%	62%	85%	23%

P7	Human	Error	P8	Human	Error	P9	Human	Error
56%	65%	9%	46%	70%	24%	44%	60%	16%
33%	40%	7%	33%	50%	17%	36%	50%	14%
44%	60%	16%	46%	60%	14%	39%	50%	11%
24%	20%	4%	60%	90%	30%	24%	40%	16%

49%	50%	1%	59%	65%	6%	41%	55%	14%
54%	65%	11%	49%	55%	6%	49%	60%	11%
47%	55%	8%	54%	70%	16%	32%	45%	13%
54%	60%	6%	64%	60%	4%	43%	60%	17%
46%	55%	9%	58%	65%	7%	46%	55%	9%
45%	50%	5%	57%	60%	3%	41%	50%	9%
53%	75%	22%	50%	90%	40%	37%	75%	38%
56%	60%	4%	53%	70%	17%	35%	60%	25%
39%	50%	11%	68%	90%	22%	45%	55%	10%
49%	60%	11%	63%	75%	12%	56%	85%	29%
55%	70%	15%	53%	70%	17%	53%	70%	17%
36%	40%	4%	64%	80%	16%	32%	50%	18%
30%	55%	25%	30%	60%	30%	57%	75%	18%
67%	70%	3%	64%	80%	16%	69%	70%	1%
50%	60%	10%	38%	65%	27%	56%	80%	24%
21%	40%	19%	35%	50%	15%	26%	40%	14%
58%	75%	17%	67%	70%	3%	64%	60%	4%
42%	60%	18%	50%	60%	10%	47%	45%	2%
42%	50%	8%	56%	50%	6%	44%	55%	11%
47%	65%	18%	63%	80%	17%	53%	60%	7%
46%	75%	29%	64%	85%	21%	52%	65%	13%
31%	50%	19%	54%	65%	11%	36%	30%	6%
52%	60%	8%	60%	75%	15%	26%	35%	9%
22%	30%	8%	56%	75%	19%	22%	30%	8%
52%	85%	33%	49%	90%	41%	55%	80%	25%
54%	60%	6%	54%	80%	26%	69%	75%	6%

22. Semantic Matching after Frequency Normalization with WUP >= 0.5

	P1	Human	Error	P2	Human	Error	P3	Human	Error
Q1	54%	70%	16%	42%	60%	18%	20%	55%	35%
Q2	33%	60%	27%	41%	70%	29%	21%	50%	29%
Q3	31%	50%	19%	39%	65%	26%	15%	40%	25%
Q4	35%	55%	20%	58%	60%	2%	35%	60%	25%
Q5	24%	40%	16%	35%	55%	20%	41%	50%	9%
Q6	24%	50%	26%	55%	50%	5%	32%	55%	23%
Q7	43%	45%	2%	54%	75%	21%	11%	35%	24%
Q8	30%	40%	10%	60%	60%	0%	30%	50%	20%
Q9	46%	50%	4%	46%	65%	19%	46%	65%	19%
Q10	36%	65%	29%	43%	60%	17%	43%	50%	7%
Q11	20%	60%	40%	40%	75%	35%	40%	70%	30%
Q12	35%	65%	30%	44%	60%	16%	53%	70%	17%

Q13	29%	60%	31%	58%	80%	22%	29%	55%	26%
Q14	41%	60%	19%	55%	60%	5%	34%	65%	31%
Q15	30%	55%	25%	53%	85%	32%	45%	60%	15%
Q16	52%	60%	8%	62%	70%	8%	21%	55%	34%
Q17	40%	50%	10%	40%	65%	25%	20%	45%	25%
Q18	36%	60%	24%	64%	70%	6%	41%	65%	24%
Q19	18%	50%	32%	27%	60%	33%	27%	55%	28%
Q20	21%	50%	29%	21%	55%	34%	35%	50%	15%
Q21	57%	75%	18%	68%	85%	17%	41%	40%	1%
Q22	32%	60%	28%	57%	70%	13%	41%	65%	24%
Q23	24%	50%	26%	36%	75%	39%	30%	65%	35%
Q24	46%	60%	14%	46%	65%	19%	31%	60%	29%
Q25	34%	65%	31%	52%	75%	23%	34%	65%	31%
Q26	38%	60%	22%	60%	65%	5%	30%	55%	25%
Q27	42%	55%	13%	42%	60%	18%	8%	35%	27%
Q28	43%	60%	17%	43%	75%	32%	22%	50%	28%
Q29	38%	70%	32%	47%	85%	38%	28%	75%	47%
Q30	45%	50%	5%	60%	70%	10%	30%	60%	30%

P4	Human	Error	P5	Huamn	Error	P6	Human	Error
47%	65%	18%	42%	55%	13%	42%	50%	8%
29%	55%	26%	33%	50%	17%	48%	75%	27%
31%	40%	9%	33%	55%	22%	49%	60%	11%
46%	50%	4%	27%	55%	28%	35%	50%	15%
49%	65%	16%	43%	65%	22%	51%	45%	6%
71%	65%	6%	58%	65%	7%	50%	60%	10%
65%	75%	10%	32%	55%	23%	43%	65%	22%
70%	70%	0%	53%	75%	22%	40%	40%	0%
69%	70%	1%	46%	65%	19%	46%	60%	14%
62%	55%	7%	43%	70%	27%	57%	75%	18%
57%	65%	8%	50%	80%	30%	50%	85%	35%
44%	55%	11%	44%	75%	31%	44%	60%	16%
55%	65%	10%	39%	70%	31%	48%	65%	17%
68%	60%	8%	55%	75%	20%	48%	65%	17%
68%	70%	2%	45%	75%	30%	45%	60%	15%
52%	60%	8%	41%	70%	29%	62%	75%	13%
57%	60%	3%	50%	80%	30%	20%	50%	30%
72%	70%	2%	57%	75%	18%	50%	60%	10%
36%	50%	14%	55%	85%	30%	46%	65%	19%
54%	60%	6%	42%	80%	38%	56%	80%	24%
70%	60%	10%	57%	65%	8%	49%	55%	6%
54%	55%	1%	41%	65%	24%	49%	65%	16%

32%	65%	33%	54%	75%	21%	50%	75%	25%
74%	75%	1%	39%	60%	21%	46%	60%	14%
66%	70%	4%	52%	75%	23%	52%	80%	28%
65%	65%	0%	38%	60%	22%	45%	55%	10%
42%	65%	23%	33%	50%	17%	25%	50%	25%
54%	70%	16%	36%	65%	29%	32%	40%	8%
63%	80%	17%	47%	85%	38%	47%	80%	33%
53%	65%	12%	60%	85%	25%	60%	85%	25%

P7	Human	Error	P8	Human	Error	P9	Human	Error
51%	65%	14%	42%	70%	28%	40%	60%	20%
33%	40%	7%	33%	50%	17%	36%	50%	14%
44%	60%	16%	46%	60%	14%	39%	50%	11%
23%	20%	3%	58%	90%	32%	23%	40%	17%
49%	50%	1%	59%	65%	6%	41%	55%	14%
53%	65%	12%	47%	55%	8%	47%	60%	13%
47%	55%	8%	54%	70%	16%	32%	45%	13%
50%	60%	10%	60%	60%	0%	40%	60%	20%
46%	55%	9%	58%	65%	7%	46%	55%	9%
45%	50%	5%	57%	60%	3%	41%	50%	9%
53%	75%	22%	50%	90%	40%	37%	75%	38%
56%	60%	4%	53%	70%	17%	35%	60%	25%
39%	50%	11%	68%	90%	22%	45%	55%	10%
48%	60%	12%	61%	75%	14%	55%	85%	30%
55%	70%	15%	53%	70%	17%	53%	70%	17%
35%	40%	5%	62%	80%	18%	31%	50%	19%
30%	55%	25%	30%	60%	30%	57%	75%	18%
67%	70%	3%	64%	80%	16%	69%	70%	1%
49%	60%	11%	36%	65%	29%	55%	80%	25%
21%	40%	19%	35%	50%	15%	26%	40%	14%
57%	75%	18%	65%	70%	5%	62%	60%	2%
41%	60%	19%	49%	60%	11%	46%	45%	1%
36%	50%	14%	48%	50%	2%	38%	55%	17%
46%	65%	19%	62%	80%	18%	51%	60%	9%
43%	75%	32%	60%	85%	25%	49%	65%	16%
30%	50%	20%	53%	65%	12%	35%	30%	5%
50%	60%	10%	58%	75%	17%	25%	35%	10%
22%	30%	8%	54%	75%	21%	22%	30%	8%

50%	85%	35%	47%	90%	43%	53%	80%	27%
53%	60%	7%	53%	80%	27%	68%	75%	7%

23. Semantic Matching after Freq. Norm. with KAON

	P1	Human	Error	P2	Human	Error	P3	Human	Error
Q1	100%	70%	30%	75%	60%	15%	25%	55%	30%
Q2	50%	60%	10%	50%	70%	20%	0%	50%	50%
Q3	50%	50%	0%	50%	65%	15%	0%	40%	40%
Q4	0%	55%	55%	0%	60%	60%	0%	60%	60%
Q5	50%	40%	10%	25%	55%	30%	50%	50%	0%
Q6	25%	50%	25%	50%	50%	0%	0%	55%	55%
Q7	67%	45%	22%	67%	75%	8%	0%	35%	35%
Q8	50%	40%	10%	50%	60%	10%	25%	50%	25%
Q9	50%	50%	0%	0%	65%	65%	50%	65%	15%
Q10	50%	65%	15%	50%	60%	10%	50%	50%	0%
Q11	50%	60%	10%	50%	75%	25%	50%	70%	20%
Q12	50%	65%	15%	50%	60%	10%	50%	70%	20%
Q13	100%	60%	40%	100%	80%	20%	0%	55%	55%
Q14	0%	60%	60%	0%	60%	60%	0%	65%	65%
Q15	50%	55%	5%	50%	85%	35%	0%	60%	60%
Q16	100%	60%	40%	50%	70%	20%	50%	55%	5%
Q17	0%	50%	50%	0%	65%	65%	0%	45%	45%
Q18	100%	60%	40%	100%	70%	30%	0%	65%	65%
Q19	50%	50%	0%	50%	60%	10%	50%	55%	5%
Q20	50%	50%	0%	17%	55%	38%	50%	50%	0%
Q21	56%	75%	19%	56%	85%	29%	0%	40%	40%
Q22	100%	60%	40%	100%	70%	30%	0%	65%	65%
Q23	50%	50%	0%	50%	75%	25%	50%	65%	15%
Q24	50%	60%	10%	50%	65%	15%	0%	60%	60%
Q25	67%	65%	2%	67%	75%	8%	33%	65%	32%
Q26	67%	60%	7%	33%	65%	32%	33%	55%	22%
Q27	50%	55%	5%	50%	60%	10%	0%	35%	35%
Q28	0%	60%	60%	0%	75%	75%	0%	50%	50%
Q29	100%	70%	30%	100%	85%	15%	0%	75%	75%
Q30	67%	50%	17%	67%	70%	3%	0%	60%	60%

P4	Human	Error	P5	Human	Error	P6	Human	Error
75%	65%	10%	100%	55%	45%	100%	50%	50%
50%	55%	5%	50%	50%	0%	100%	75%	25%
100%	40%	60%	50%	55%	5%	100%	60%	40%

0%	50%	50%	0%	55%	55%	0%	50%	50%
50%	65%	15%	75%	65%	10%	75%	45%	30%
50%	65%	15%	75%	65%	10%	75%	60%	15%
100%	75%	25%	67%	55%	12%	67%	65%	2%
75%	70%	5%	100%	75%	25%	50%	40%	10%
50%	70%	20%	50%	65%	15%	50%	60%	10%
50%	55%	5%	100%	70%	30%	100%	75%	25%
50%	65%	15%	100%	80%	20%	100%	85%	15%
50%	55%	5%	100%	75%	25%	100%	60%	40%
100%	65%	35%	100%	70%	30%	100%	65%	35%
0%	60%	60%	0%	75%	75%	0%	65%	65%
50%	70%	20%	50%	75%	25%	50%	60%	10%
50%	60%	10%	100%	70%	30%	100%	75%	25%
0%	60%	60%	0%	80%	80%	0%	50%	50%
100%	70%	30%	100%	75%	25%	100%	60%	40%
50%	50%	0%	100%	85%	15%	100%	65%	35%
50%	60%	10%	83%	80%	3%	83%	80%	3%
56%	60%	4%	56%	65%	9%	56%	55%	1%
100%	55%	45%	100%	65%	35%	100%	65%	35%
50%	65%	15%	100%	75%	25%	100%	75%	25%
100%	75%	25%	50%	60%	10%	50%	60%	10%
67%	70%	3%	100%	75%	25%	100%	80%	20%
67%	65%	2%	67%	60%	7%	67%	55%	12%
50%	65%	15%	50%	50%	0%	100%	50%	50%
0%	70%	70%	0%	65%	65%	0%	40%	40%
100%	80%	20%	100%	85%	15%	100%	80%	20%
67%	65%	2%	67%	85%	18%	100%	85%	15%

P7	Human	Error	P8	Human	Error	P9	Human	Error
75%	65%	10%	75%	70%	5%	75%	60%	15%
100%	40%	60%	50%	50%	0%	100%	50%	50%
50%	60%	10%	50%	60%	10%	50%	50%	0%
0%	20%	20%	0%	90%	90%	0%	40%	40%
50%	50%	0%	75%	65%	10%	50%	55%	5%
75%	65%	10%	25%	55%	30%	50%	60%	10%
67%	55%	12%	67%	70%	3%	67%	45%	22%
50%	60%	10%	50%	60%	10%	25%	60%	35%
50%	55%	5%	50%	65%	15%	50%	55%	5%
50%	50%	0%	100%	60%	40%	50%	50%	0%
50%	75%	25%	100%	90%	10%	50%	75%	25%
50%	60%	10%	100%	70%	30%	50%	60%	10%
100%	50%	50%	100%	90%	10%	100%	55%	45%
0%	60%	60%	0%	75%	75%	0%	85%	85%
50%	70%	20%	50%	70%	20%	100%	70%	30%

50%	40%	10%	50%	80%	30%	50%	50%	0%
0%	55%	55%	0%	60%	60%	0%	75%	75%
100%	70%	30%	100%	80%	20%	100%	70%	30%
50%	60%	10%	100%	65%	35%	50%	80%	30%
17%	40%	23%	50%	50%	0%	17%	40%	23%
56%	75%	19%	56%	70%	14%	56%	60%	4%
100%	60%	40%	100%	60%	40%	100%	45%	55%
50%	50%	0%	100%	50%	50%	50%	55%	5%
50%	65%	15%	50%	80%	30%	50%	60%	10%
67%	75%	8%	100%	85%	15%	67%	65%	2%
33%	50%	17%	33%	65%	32%	33%	30%	3%
100%	60%	40%	50%	75%	25%	100%	35%	65%
0%	30%	30%	0%	75%	75%	0%	30%	30%
100%	85%	15%	100%	90%	10%	100%	80%	20%
100%	60%	40%	67%	80%	13%	100%	75%	25%