Building Ontology-based Metadata to Support Arabic Quality Assurance Documentation

بناء بيانات وصفية قائمة على الأنطولوجيا لدعم وثائق ضمان الجودة العربية

Prepared by

Faisal Mohamed Alromaihi

Supervisor

Prof. Ahmad Kayed

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Computer Science

Department of Computer Science

Faculty of Information Technology

Middle East University

May-2017

AUTHORIZATION STATEMENT

I, Faisal Mohamed Alromaihi, authorize Middle East University (MEU) to provide copies of my thesis to the concerned libraries, establishments, and institutions upon request.

Name: Faisal Mohamed Alromaihi

Date: 31-5-2017

Signature:

اقرار تفويض

انا فيصل محمد الرميحي افوض جامعة الشرق الاوسط بتزويد نسخ من رسالتي للمكتبات المعنية، المؤسسات، الهيئات عند طلبها.

الاسم: فيصل محمد الرميحي

التاريخ: 2017-5-31

التوقيع:

Examination Committee Decision

This is to certify that the thesis entitled "Building Ontology-based Metadata to Support Arabic Quality Assurance" was successfully defended and approved on 31-05-2017.

Examination Committee Members	**Signature**

(Supervisor)

Prof. Ahmad K. A. Kayed

Professor

Dean Faculty of IT

Middle East University

(Internal Committee Member)

Prof. Hamza A. A. Alsewadi

Professor

Middle East University

(External Committee Member)

Prof. Abdelfatah A. Tamimi

Professor

Alzaytoonah University

Acknowledgment

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

Special thanks to my parents for all the moral support and the amazing chances they have given me over the years.

Table of Contents

List of Tables

List of Figures

List of Abbreviations:

AGLS	Australian Governments Locator Service
API	Application Programming Interface
AS	Australian Standard
AWN	Arabic WordNet
C-value	Collocation-value
DC	Dublin Core
GSD	Geospatial Data
ISO	International Organization of Standardization
MML	Medical Markup Language
MSE	Mean Square Error
NTDB	National Topographic Data Base
OI-models	Ontology-Instance Models
QA	Quality Assurance
QAD	Quality Assurance Documents
RDF	Resource Description Framework
RKMS	Record keeping Metadata Scheme
SGML	Standard Generalized Markup Language
SPIRT	Strategic Partnership with Industry – Research & Training
TFIDF	Term Frequency Inverse Document Frequency
VERS	Victorian Electronic Records Strategy
WuP	Wu and Palmer
XML	Extensible Markup Language
XSLT	eXtensible Style Language Transformations

Building Ontology-based Metadata to Support Arabic Quality Assurance Documentation

Prepared by

Faisal Mohamed Alromaihi

Supervisor

Prof. Ahmad Kayed

Abstract

The ontology has a vital role in knowledge management especially for quality assurance department that is based on knowledge abundance. The primary challenge to build ontology for such big database associated with concept extraction.

This work aims to use a semi-automatic approach using KAON tool Text-To-Onto framework in purpose to extract the proper concepts for the aggregation of Arabic quality assurance documents based on its metadata by using automatic translation tool (Google translate), to translate Arabic documents into an English, so that we can process the documents through the proposed tool.

This study will utilize the experimental methodology and be calculating the error rate in automatic concepts extraction compared with the concepts semi-manually extracted. The result shows that Text-To-Onto and automatic translation tool (Google translate) framework is considered valuable and efficient approach with maximum error rate 14% and highly sufficient performance with min error rate equal to 6%, which means a success rate of 94%, furthermore the minimum number of relevant concepts retrieved are 69%; also the maximum number of relevant concepts retrieved are 84%. Which is very good for especially in Arabic ontology building. also the experiments

show the sensitivity of error rate to the data size where the error rate decreases gradually with increasing data size and after particular size, the error rate shows data size independent behaviors, the research recommends to probability build an ontology for quality assurance department based on the concept extracted in.

Keywords: Ontology, Arabic ontology, Quality assurance, Concept extraction ,KAON, Metadata, Google translate.

بناء بيانات وصفية قائمة على الأنطولوجيا لدعم وثائق ضمان الجودة العربية

إعداد

فيصل محمد الرميحي

إشراف

الأستاذ الدكتور أحمد الكايد

المُلخص

تتخذ الانتولوجي دور فاعل ورئيس في إدارة المعلومات، لا سيما في أعمال قسم ضمان الجودة المعتمدة على وفرة البيانات والمعلومات، ومن التحديات الرئيسية لبناء انطولوجيا لمثل هذا الحجم من قواعد البيانات يتمثل في استخراج المفاهيم الاساسية المحددة لجميع الملفات.

جاءت هذه الدراسة بهدف استخدام طريقة شبه أتوماتيكية من خلال توظيف برنامج Onto – TEXT-To أحد أدوات حزم برامج KAON لاستخراج المفاهيم الأساسية لملفات قسم ضمان الجودة باللغة العربية التي من شأنها تسهيل بناء الانطولوجيا تدعم لقواعد بيانات ضمان الجودة باللغة العربية. وذلك بمساعدة مترجم آلي (Google translate) لترجمة الوثائق من اللغة العربية الى الإنجليزية ,ليتسنى لنا معالجة الوثائق بالأداة المقترحة.

وظفت الدراسة المنهج التجريبي في الدراسة من خلال القيام بعدة تجارب عملية لاستخراج المفاهيم من خلال البرنامج TEXT-To-Onto ومقارنة هذه النتائج مع المفاهيم الأساسية المستخرجة يدوياً من الملفات الخاضعة للتجربة وحساب نسبة الخطأ. وقد أظهرت النتائج أن البرنامج Text-To-Onto والمترجم الألى (Google translate) المستخدم يعتبر برنامج فاعل

وقيم باعتبار أن أعلى نسبة خطأ سجلت كانت 14% وأقل نسبة خطأ سجلت 6% ما يعني نسبة نجاح بمقدار 94% , اضف على ذلك فإن الحد الأدنى لعدد المفاهيم ذات الصلة التي تم استخراجها هو 69% ؛ أيضا الحد الأقصى لعدد المفاهيم ذات الصلة المستخرجة هي 84%.

وتعتبر هذه النسبة جيدة جداً لبناء أنطولوجيا باللغة العربية، وقد أظهرت الدراسة أثر حجم البيانات في نسبة الخطأ فكانت زيادة حجم البيانات يقابلها نقصان في نسبة الخطأ حتى حجم معين ومن ثم أظهرت النتائج نسب خطأ غير متأثرة او متغيرة بحجم البيانات، وقد أوصت الدراسة بإمكانية بناء انطولوجيا لقسم ضمان الجودة بالاعتماد على المفاهيم المستخرجة فيه.

الكلمات المفتاحية: الانتولوجي، الانتولوجي العربي، ضمان الجودة، استخراج المفاهيم ، كايون، البيانات الوصفية ، مترجم جوجل.

CHAPTER ONE

Introduction

Overview

Chapter one has explored the research theoretical background including metadata, quality assurance, ontology and its related concept extracting issue. Also, this chapter stated the research question and objective through highlighted research contributed problem, and addressing the research methodology.

1.1 Introduction

Recently, Ontologies occupy a large attention especially in the information systems field by intelligent researches, where the ontology involved in its various aspects such as web technologies, database integration, multi agent system, natural language processing and so forth. (Roussey et.al, 2011)

Ontology concerned in reality conceptions investigation and its nature, through define the reality concept properties, objects and relations in each area, where ontology proposed any existing domain in hierarchy fashion based on its dominant concepts, it is properties and relations. However, the main objective of ontologies is supporting knowledge sharing among people and projects in view of enhancing reusability, interoperability, enhanced consistency of research results, and navigation improvement. (Shaheen et. al., 2017)

Organizational quality domain is considered one of the various ontology domains, organizational quality took its significance since its necessity in our societies for both customers and the organization itself. Quality is an indicator that represents trust, credibility, worthy and reliability in organization from customers view, thus for it considered as an important factor affects customer's attitudes towards the organization. In other hand, quality is crucial aspect of the organization's business. Quality affects customer loyalty and satisfaction, risks and cost reduction, earning accreditation and

organization reputation. Moreover, The quality decision based on the abundances of evidences, that are based on finding document and data within large database, thus the consistency and compatibility of research results (data mining) is considered avital point in quality domain (Manghani , 2011).

1.1.1 Metadata

Metadata concept is required by data mining, which helps quality experts to find and know about data stored, metadata is a data of data (Bui, 2012). It has three types, descriptive, administrative and structural. Descriptive metadata is used to describe the information object, in terms of what the object is and/or what it contains. Administrative metadata concerns with information object creation and preservation, who, what, where aspects. Finally, the structural metadata concerned with the relationships and the associations between individual information objects (Sistrunk, 2002), it represents elements such that subject, author, date etc. Structural metadata typically depends on both organizational and system needs requirements, and uses standard schemas representation such as Dublin Core (DC). However, administrative and descriptive metadata are created based on a specific content (Benson, 2012).

Metadata is used to support different kinds of operations. The purpose typically when dealing with metadata is to find and locate the information resource of interest (recourse discovery) based on specific criteria. The user then should be allowed to evaluate the relevancy of resource retrieval (Johnston, 2005). The more relevant the information is; the higher value of metadata (documentation) quality. Retrieval relevancy is a good evidence quality assurance for metadata.

Metadata can be expressed in a wide range of human and artificial languages and formats. However, metadata creators often generate and organize metadata without

concerning about the requirements of the functions or quires to be performed over it. This is a general semantic problem when processing and reacting on metadata. Hence, many researches have proposed and developed metadata as a mechanism for expressing the "semantics" of information, to enhance information retrieving and understanding. In consequence, merging metadata languages with ontology can be considered as the basis for a new framework of information management (Sicilia, 2006).

1.1.2 Quality assurance

It responsible for supplying evidences to quality management; these evidences show confidently that the quality requirements are fulfilled. Quality experts face many challenges such as catching proper data in view of emission source through the large amount of documents and data, time allocated to checks emission source, archiving information and data, and quality confidentiality checks of source information (IPCC, 2000).

Quality analysts, who look beyond the secrets of information, took advantages of metadata to serve the documentation quality of their business purposes. Metadata can be read comprehensively in both technical and business focus, where technical metadata refers to data movement, translation, aggregation and presentation of data, however business metadata highlights the defined information means to business, thus metadata describes the contents of data stored, leading to enhancement in analyst understanding of the data being used (Parankusham & Madupu, 2006).

Hence, Metadata can be designed and formulated with different schemes or structures. Metadata schema represents a set of metadata elements (metadata semantics) that is designed for a specific purpose, where these elements or semantics have names and values that can be represented in different formats. Optionally, metadata schemes

define the rules for content formulation and representation, for example, using controlled vocabulary or capitalization rules. They may also use syntax rules that determine how the content of metadata should be encoded. However, metadata schemes that don't subject to a prescribed syntax rules is called syntax independent metadata scheme. There are many definable syntax that metadata could be encoded with, such as Standard Generalized Markup Language (SGML) and Extensible Markup Language (XML) (NISCO press, 2004).

1.1.3 Ontology

To provide this efficient browsing for knowledge, the user have to be powered with concepts and relationships associated with specific subject areas; this could be achieved by producing metadata-ontology. Ontology is a model for knowledge in a given domain (such as medicine, agriculture, etc.) that consists of components like concepts (classes or objects), attributes (traits of objects), relations (connections between objects), and instances (the actual data of objects). Figure 1-1 shows an example of an ontology expressing these components (Sini, Salokhe, Pardy, Albert, Keizer & Katz, 2007).
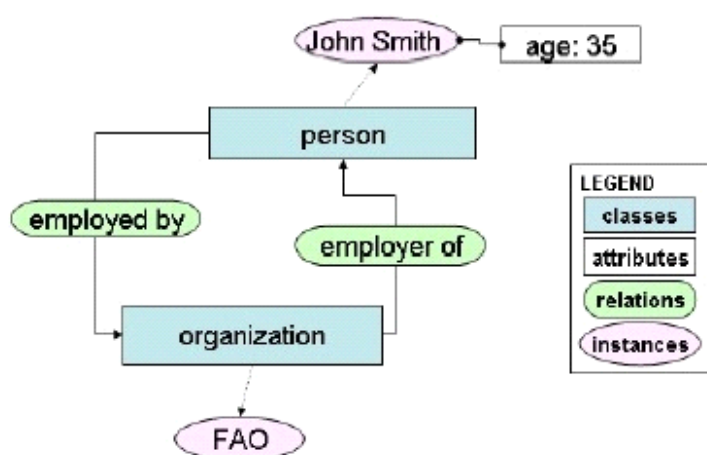
Figure 1-1:Example of an ontology expressing relationships between persons and organizations (Sini, Salokhe, Pardy, Albert, Keizer & Katz, 2007).

1.1.4 Concept extraction

In organizations with large document collections such as digital libraries, selecting only the requested information (not the whole document) is very important mechanism in retrieving, so using keywords to represent documents is a promising method that would help in classifying documents as relevant or not. However, determining which concepts are associated to each document is not an easy task, hence, various concept extraction methods have been released to extract concepts that best describe documents (Abascal-Mena & Rumpler, 2007).

1.1.5 KAON

In this study, a quality comparison between Ontology-based metadata and a semi-manual metadata will be performed in the domain of Quality Assurance Documents (QAD), to show if metadata that depends on ontology concepts could be a better structure for storing and retrieving, than metadata stored in the ordinary format. Metadata comparison and evaluation will be performed for Arabic language documents, in terms of retrieval relevancy.

The Karlsruhe Ontology and semantic web (KAON) framework will be used to extract ontology concepts from documents. KAON has been used to develop several web applications that is semantic based, it provides tools for discovering, managing, presenting ontologies and metadata (Bozak et al., 2002).

1.1.6 WordNet

WordNet is a computational reflection for dictionaries in real life, where the WordNet functionality as same as dictionary in giving meaning for language concepts and semantic relation that link synonym sets. However, the WordNet perform a lexicographic information database support machine language and application based

Concepts and semantics such as ontology, data mining application and so forth. (Miller, 1995)

In this area, area of Natural language processing (NLP) the WordNet became a repository of lexical semantic information useful for ontology building and various knowledge − based applications, the most popular WordNet which is built for English language "Princeton's English WordNet", is based on last mentioned success, a lot of effort have been supporting other languages WordNet, such as Euro WordNet.(Alkalifa, Rodriguez, 2010)

The Arabic WordNet is still in the construction phases, the importance of having such WordNet that Arabic language is the official language for hundreds of millions in twenty middle east and north African countries, and the Islam language among various ethnicities in the world.(Rodríguez et al.,2007)

1.1.7 Protégé

Protégé is programming software based on Java programming language perform an open-source platform running with extendable architecture in various operating system as a standalone system or as java library. (Jurˇc´ık , 2010)

The core of protégé environment simplify in building and editing ontologies that download and used by others, its support both Web Ontology Language OWL- 2 and Resource description framework (RDF) as World Wide Web Consortium (W3C) specification. The most powerful character of protégé represents in user friendly and orders simplification, it also supported ontology visualization through its plug-in that has various functionality. However, protégé offers ontology error checks or unspecified relations, through running reasoning statement which detect the classes or relation

inconsistence that could harm the semantic properties of classes and relations.(Hokstad, 2015)

In this study, first we attempted to use the Portege software because it can create and display ontology in Arabic also manage conceptual terminology in ontology, but there are such barriers through using Protégé. The difficulty of extracting concept and inaccuracy in the software results.

Based on previous barriers shown in Protégé software, we decided to change the software come over these barriers. The new software found defined KAON that offers the automatic concept extraction service through Text-To-Onto tool.

1.2   Problem Statement

Extracting and creating ontologies have been widely studied for English language, such as the English linguistic ontology named "WordNet", in other hand, research on the Arabic ontologies were rarely recognized in the literature . To cope this challenge, the research aim to use the translation tools to exploit the ontology building software in the Arabic applications, through translating the Arabic corpus data to the destination language supported via ontology software's such as English, which is the destination language of the research.

The main idea of translation was delivered based on the existence fact that ontology represent vocabulary even if it is specialized to the certain domain, to be more precise, vocabulary does not qualify for the ontology, but the conceptualizations do which is the intended captures, thus the term translating process in ontology does not amend ontology conceptually.

Hence, this research aims to build an ontology-based metadata supported Arabic Language, the focuses will be to build an Arabic metadata basis which contributes with concept extraction process, the experiment domain will cover the quality assurance database since its importance for all organizations, so it could be very helpful to create such metadata to provide an efficient search and retrieval for their documents. And since documentation and metadata creation for this domain is rarely used in literature; it is useful to be investigated over.

This study aims to answer the following main questions:

- How ontology-concepts could be used to support building metadata for Arabic documents?
- How metadata could be built from Arabic documents?
- How English-based tool could be used to serve Arabic documents?

The above questions will be measured by testing the following hypothesis:

Hypothesis: **The quality of the proposed method of building Arabic ontology-based metadata using an English translation tool, will outperform the existing semi-manual metadata.**

1.3   Research Objective

Based on the previous mentioned problem, the objective of this research is defined by the following points.

- To build metadata that would produce high document retrieval precision, and fast searching for QAD.

- To show the role of using ontology extraction to create a metadata format that would enhance document retrieving in QAD compared to the semi-manual metadata formats.

- To investigate using the proposed method for developing ontology-based metadata for Arabic Language documents.

- To promote and show the role of using tools in extracting and creating ontology concepts.

1.4   Research significance

The research significance can be formulated in:

- Building metadata that would enhance searching and retrieving unstructured documents, using ontology extraction and creation.

- Supporting Arabic language documentation by presenting a new method for creating automatically an ontology-based metadata, since Arabic language is rarely supported by various ontology creation tools.

- Utilizing Translation tool to overcome the weakness in ontologies based Arabic building Software.

- Providing and evaluating metadata for a new uninvestigated domain of documentation, which is documentation in the Middle East University QAD.

1.5   Methodology

This research will utilize an experimental methodology to address research objectives. A set of comparative experiments will be conducted between two metadata: the semi-manually generated metadata and the ontology-based metadata using a translation tool. A set of Arabic text documents will be used and applied for the two methods, and a set of queries will be performed over them. Their outputs will be judged by quality assurance experts, and based on their judgments, precision will be computed for each query performed over both metadata. All quantitative results are then will be used to evaluate the two methods, answer the research questions, and prove the hypothesis. Comparing the quantitative results for the two methods will determine their retrieval quality, and determine which one outperformed the other. The Procedure of the proposed concepts extraction method will peruse the following steps:

- **Arabic data (text document) collection** from corpuses used in QA department of a university.

- **Building semi-manual metadata** for these Arabic text documents.

- **Document translation** using Automatic translation tool (Google), from Arabic to English language.

- **Ontology concepts extraction** from the English text documents, using KAON environment.

- **Retranslate the extracting concepts to Arabic language** again using Google translate.

- **Comparison** over both extracted concepts**.**

- **Results evaluation** for quality (precision) of concepts extraction.

- **Results summarization and representation.**

1.6   Organization of the Thesis

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

Chapter 2 discusses a theoretical background and literature.

Chapter 3 introduces the methodology of this research. The research methodology has the following main steps: Data collection, Building semi-manual metadata, Document translation, Ontology concepts extraction, Retranslate, Comparison, Results evaluation.

.Chapter 4 explains the experimental results and the evaluation.

Chapter 5 presents conclusion of this thesis and future work.

CHAPTER TWO

Literature Review & Related Works

Overview

Literature has explored many various previous works on ontology concepts extraction and metadata building; some of these works are represented below.

2.1   Related Work

Mckemmish & Acland (1999) proposed an Australian project named as Strategic Partnership with Industry – Research & Training (SPIRT) project, which focuses on record-keeping metadata after doing a lot of business analysis, organizational and social context of record-keeping , record-keeping sector specific metadata schemes and development of Dublin Core (DC) with Australian Governments Locator Service (AGLS) metadata sets. The AGLS metadata is based on DC metadata shared information source with entire levels of aggregation such as agencies or government services. This project is formulated to support the persistence in networked environments of reliable, authentic, meaningful and accessible by defining high level set of metadata. Recordkeeping metadata framework provided by people react with each other in organization relation performs business by creating and managing business records in addition to retrieve and recall these records. So the frame work based into various entities categorized as Business entities, people entities and records entities. All these entities have an inter-relationship connections need unique Identifies and standardized descriptive metadata. High level metadata in this project is established by implementation several standards consist of Australian Standard (AS) 4390 Records management, University of Pittsburg functional requirement, Electronics records developed at university of British Columbia, DoD5015.2 STD design criteria standard, Metadata model for university of Pittsburg, ISAD, EAD , Australian Registry and series system, and the Victorian Electronic Records Strategy (VERS) metadata set. Record

Metadata scheme represented in three classes based on system entities, Business Metadata, Agent Metadata, and Record Metadata.

The SPIRT Record Keeping Metadata Scheme (RKMS) modeled in Resource description framework (RDF) and is expressed in XML, the significant features of this metadata scheme is scalability and expression complex relationships. Mapping the RKMS with existed DC-AGLS scheme shows direct match in most simple cases in addition to misaligned in complicated case especially in relation to the type of elements in AGLS and DC set, where the RKMS shows clear relation and position of elements, However, RKMS is more extensive access and uses metadata elements relating to whole managing access and use terms and condition aspects. It defined and described great features in business context linked to people and agents.

Araki, Ohashi, Yamazaki, Hirose, Yamashita, Yamamoto, Minagawa, Sakamoto & Yoshihara (2000) have developed Medical Markup Language (MML) to store, access and exchange medical data within any physical worldwide location, and specifically in Japan. They used XML as meta-language for MML, such that each document is characterized by XML module structure to enhance linking between documents. For querying and replying, additional data exchange specification was used. The XML-based MML Language consists of two parts, the header and the body. The header is for data transmission, and the body contains the module items, in which each module contains document information and module content. By developing this XML-based language, many medical module contents are defined now, such as health insurance information, basic clinic information, diagnosis information and many others. The use of XML as meta-language for MML has led to the enhancement of data exchange of this language.

Teng (2000) employed XML to investigate if it is possible for geospatial metadata to be specified and implemented using XML standards, and investigated how we could make fast search for data queries in large collections of geospatial data, by using a specific way of data indexing which combines keywords and date ranges with spatial ranges. He translated metadata files of Canadian National Topographic Data Base (NTDB) into International Organization of Standardization (ISO) XML metadata files using eXtensible Style Language Transformations (XSLT) and a specific Java-program for this purpose. Comparison experiments were run over thousands of ISO XML metadata files between two search engines: the geographical data index approach (Geospatial Data (GSD) index), and the relational database approach. The results showed that the proposed index is more efficient than the other, it is faster by 2.5 times.

Widener, Eisenhauer & Schwan (2002) developed a novel approach to provide high performance data transmission for structured data, by decomposing it into binary low level data encoding, while also maintaining the data structures readable by users. The process of decomposing is represented by three separate steps: metadata discovery for the message being transmitted, the binding between metadata format and a particular data unit using Application Programming Interface (API) call, and finally the process of message transmission by translating data messages from in-memory format to a wire format. The proposed approach was implemented using XMIT (XML Metadata Integration Toolkit) tools, in which data messages with XML formats were used being transformed using XML schema specification datatypes. A set of experiments were conducted and showed high performance in transmitting binary data by XMIT, compared to naïve structured data exchange methods. Results indicated that we can get

large gains in usability and performance by using XML as a metadata definition language.

Potok, Elmore, Reed & Samatova (2002) developed a process for converting HTML ontology of Internet newspapers into an XML, using intelligent software agents. These agents must be able to understand the given ontology and automatically get the required information from websites. Each newspaper site has its own ontology that can be represented with graphs to characterize their elements. RDF was used for ontology representation, and software retrieval agents were used to query and retrieve information from newspapers websites. These agents responsible to parse pages, interpret RDF ontology, retrieve relevant articles and convert them to XML. The proposed ontology-based conversion process has been tested on more than 17 newspapers sites, and showed a quite well work. It had been operated within the Virtual Information Processing Agent Research (VIPAR) at the US Pacific Commands Virtual Information Center.

Lee (2003) introduced metadata aspects in a framework called Context Interchange (COIN) as a context interchange platform, highlighted the aspects of metadata representation and the management layer in the context interchange framework. The main concept missing in the previous COIN platform is the knowledge sharing and reuse easily. COIN is a bridge system built to link the contextual aligned between heterogeneous resources of data. In conceptual level COIN model includes ontological information, contextual information and source information, which collaboratively represent context mediation. COIN data lifecycle illustrated in: collection, storage, processing, sharing and reuse stage. The mediation engine extended-COIN, or eCOIN integrated by reusing metadata based on XML representation,

provides user interface for eCION running application, and create a managing metadata registry. Each component of COIN model connected to metadata layer internally, with respect to metadata representation, researcher took advantage of XML based representation. These XML metadata are converted to RDF format and also the RDF format translate into Rule Markup Language (RuleML), Relational-Functional Markup Language (RFML) and Prolog (a logic programming language) based on XSLT transformation series. Prolog is context mediation consumed by eCOIN. By this model the user interface was created, and registry was setup to managing and accessing metadata files, indicates accomplishment of objectives study, the main limitation of this project is the small number of system users which raveled the concurrent file access evaluation.

Motik, Maedche & Volz (2005) presented an approach for ontology model of representation and querying that try to realize the requirements of business wide applications. They argued that the existing approaches of conceptual modeling do not support important features for enterprise systems. The proposed ontology structure provides a way for conceptual querying in a hope of improving information extraction within large document bases. Well-established technologies were used for evaluating queries, such as relational databases, and meta-concepts were modeled using well-defined semantics. Ontology modeling and querying was implemented within KAON tool suite, and showed rich capabilities and features for information browsing and navigation over other existing approaches.

Aktas (2006) introduced a design based on XML metadata services to address some problems associated with metadata requirements, with a goal to design semantics for XML metadata services. This design was tested over two web service standards, the

Universal Description, Discovery, and Integration (UDDI) and Web Services Context (WS-Context), and examined XML metadata services for its ability to provide the proper knowledge and information. Extended UDDI and hybrid WS-Context XML Metadata Services were implemented successfully with various information grids using the proposed prototype, and the results showed that services provided by WS-Context XML Metadata outperformed the performance of UDDI services by 30%.

Abascal-Mena and Rumpler (2007) proposed a semantic-based approach to enhance document retrieval, by using concepts extracted from documents themselves as metadata tags to annotate documents. Their work was deployed in a digital library context, in Consultation of entire text versions of theses (CITHER) of National Institute of Applied Sciences (INSA), Lyon. Concepts from digital theses were extracted to build ontology of new terms and relationships, with the intent to add new semantic metadata for these theses. The proposed approach showed two advantages in information retrieving, first, the use of semantic annotations allow exploiting the whole content of digital theses, and second, it can provide users with different alternatives for their access requests. The construction of ontology in the proposed approach was the first step to enhance information access in digital libraries.

Sini et al. (2007) developed an ontology using resource description framework (RDF) with KAON application programming interface, and using articles of metadata available from cataloging and indexing systems of The Food and Agriculture Organization (FAO) of the United Nations. The aim of building this ontology was to provide more efficient and effective knowledge browsing and extraction from food and nutrition publications. A knowledge model of concepts and their relationships was built from two different bibliographic databases, in which they were merged and unified to

provide better access for users. The produced metadata-ontology has provided many functions, such as easy navigation to articles by following semantic links, controlled vocabulary for user's query terms, cross-linguistic information retrieval, and other much functionality.

Cjien-cheng and Chao-chen (2009) proposed a system design to perform metadata storage and exchange through an embedded metadata framework (EMF) structure. EMF is intended to provide metadata standards for digital images, and to provide an information exchange frame among different platforms. It is based on XML structure of XMP (Extensible Metadata Platform), which is designed for embedded metadata exchange of digital files. The proposed design combines image files and metadata; it suggests that when digital images are stored in metadata, they can be processed and retrieved in a location that is independent of their store location, such that the editing units are different from the preserving ones. When the digital files are downloaded for use, they can be shared and exchanged through Metadata XML tools. The EMF structure has some flaws; however, XMP application is considered to be a good way to combine metadata with digital files, and it could be feasible for some possible problems of embedded metadata**.**

Olfat (2013) innovated a conceptual framework to come over current spatial metadata management approach, which is separated from spatial data life cycle, inefficient and ineffective for automatic metadata content updating in synchronization with dataset modification, and not supported to end users interactivity in order to create and enhance the content of metadata. The spatial metadata lifecycle illustrated in: collection, creation, storage, publication, discovery, retrieve and access and updates. The new framework considered the integration capability in metadata creation in spatial

metadata life cycle in addition to updating synchronized with data modifications and end users involving. The new framework took advantage of Doblin core standard for storing and bunching spatial data and metadata in view of Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM) and ISO 19115 standard. Also the new framework facilitated end users interaction and involvement by utilized Web 2.0 features in order to establishing and enhancing the content of metadata. This innovated frame work can't contribute with ambiguity and heterogeneity of user generated search wards, and spatial data discovery through visualizing datasets**.**

Ahmed (2014) introduced an ontology in the Renewable energy (RE) domain to automate the renewable environment where the appropriate service provider can easily discovered in the customer researches, the proposed ontology based on three main steps, extracting RE provider Instances, extracting Customer requirement instances, matching two extracted instances using semantic similarity measurement, where the KAON software was utilized for instances extraction process. The instances defined through collected consumer inquiry documents, reports and documents of RE provides. In parallel the semi-manual approach will conduct, the evaluation phase based on the comparison between the two results, the traditional and proposed approaches result, where the result indicates 10% reduction of the error in the result demonstration.

Ribeiro, M. (2014) introduced a framework for implementing extraction non – taxonomic relation unsupervised approaches, in specific ontology and corpus domains, in purpose to enrich the ontology via non-taxonomic relation, in order to reduce the human effort in building ontologies, and provide automatic and semi-automatic

approaches support construction ontology. The non-taxonomic relation consider a vital difficult and challenges faced knowledge engineers.

Al-Hassan and AlHassan (2016). Introduced an ontology building of data breaches threats for cloud computing based on the concept extraction, discovering all concept related to data breaches threat domain, the researchers utilized KAON tool for extracting concept from huge data collection, and count on field experts to refined the concept extraction and relation.

Kontokostas, Mader, Dirschl, Eck, Leuthold, Lehmann, & Hellmann (2016) introduced an enrichment development for data life cycle in JURION platform in main two phases : quality tools integrated phase, and improving data enrichment phase by, checking external resource before importing them to the platform. JURION life cycle represents data extraction, storage, authoring, interlinking, enrichment, quality analysis, repair and publication. The development team concentrates on enhancing data quality and repair process by integrating RDF tools in existing platform, and creating a link validation service in purpose of cleaning up external metadata stored in a database platform. The developed structure utilizes RDF units for monitoring data stored in order to implement quality assurance side. The development represents the integration of RDF unit with JU units (Java framework testing unit) where JU unit enables repeating test execution in development workflow, each test done by RDF units tags mapped to individual JU unit test and report which leads to development platform tool stacks.

The Second development related to linked external metadata from pool party projects where evaluate differ methodology so supported tools will be required to check validity of links. While this is certainly useful, it is required to change the current configuration to redeployment of Pool Party. So the suggestion is to limit link lookup,

and resolved URLs by either SPARQL or Java algorithm. In this development coverage report need to be tested in respect to generation test analytic, also the internal link need to be supported by advanced configuration for end users.

Aldiery, (2017) aims to study the semantic similarity measures exploit Arabic WordNet database, several measures have been applied on Arabic WordNet (AWN) derived from various categories such as path-based measures, information content-based, feature-based measures, and hybrid measures in order to define the best performance measures on Arabic WordNet. The results illustrated that the Wup measure has a minimum mean square error (MSE) value and a highest correlation with the human rating, in other hand path-based measures has the worst performance.

2.2   Software used in the work

2.2.1 KAON Software Tool

KAON is one of the popular tools utilized for assisting ontology building, the main features and reasons to use this tool conclude in: (Ahmed, 2014)

- It is open- source tool.

- Easy to create and edit the ontology.

- Building ontology based on the underlying textual data.

- Friendly use.

Gabel et al. (2004) described the important features of KAON in the following:

- Meta Modelling which means that the concept and property are meta-concepts' instance.

- Evaluation Strategies were KAON API (ontological interface which representing different pieces of ontology such as concept, relation, attribute) checks each change occurred on ontology to avoiding inconsistent state.

- Change Notification were Ontology-Instance Models (OI–Models) interface allows receiving model updates notification.

- Lexical Layer provides ability to add lexical information such as labels or documentation on the OI-Model.

- Modularization through KAON API, KAON can build modular ontologies for same type ontologies as ontology inclusion as a link not a copy.

KAON provides middle ware layer through KAON API, and offers external servers access though KAON server. (Bozak et al. ,2002)

KAON API considers as vital and main point characterized the KAON Frame work architecture, KAON API provides a representation of objects respect the ontology parts, the objects are important for queries making, and providing classes such as Concepts, Relation, and Attribute (Instance).(Gabel et al. , 2004) Thus KAON API defined as a Model-view –controller, it provides underlying ontology consistency. Furthermore, KAON Server which is responsible for ensuring persistence, Transactional and secure multiple access through RDF repository. (Bozak et al. ,2002)

KAON distinguished on three layer named the data and remote service layer, the middleware layer, and the application layer as shown in figure 2-1, where illustrates following: (Bozak et al. ,2002)
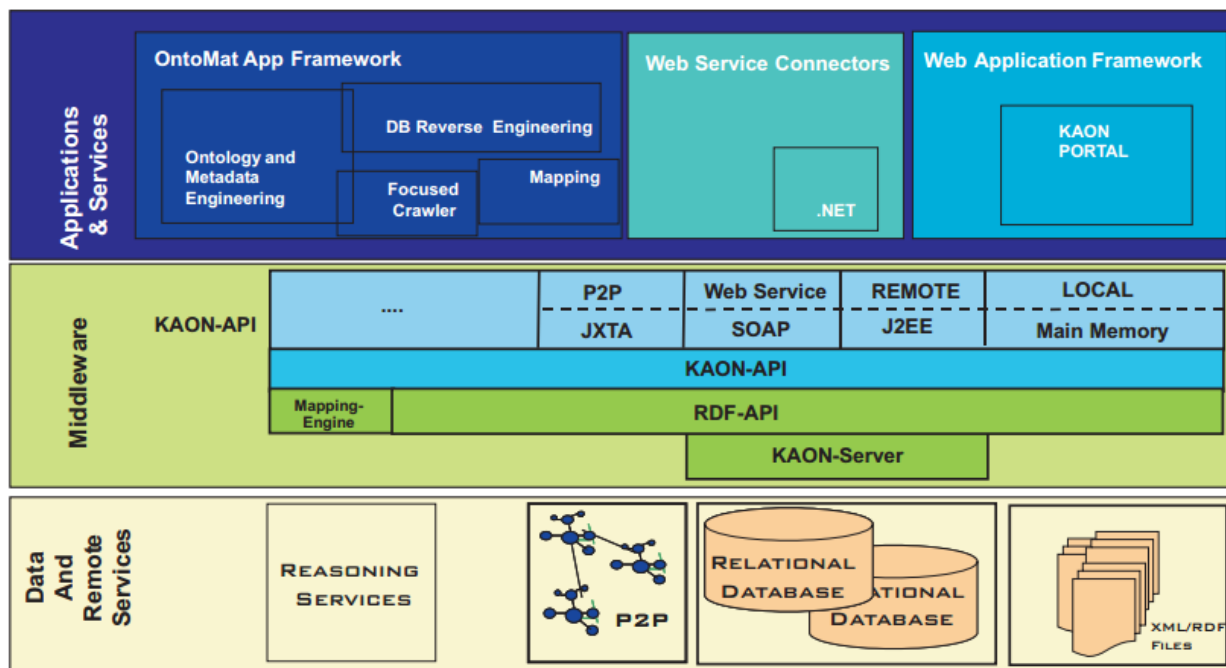
Figure 2-1:KAON Layers

Application and service layers, the clients of this layer can be either the Java – application based Onto Mat application framework Components, or the extended application of the web based KAON portal and website management frame work, where these clients contact with next layer (middleware layer) through KAON API (ontological interface).

Middleware Layer that mainly contributes abstraction ontology access that implemented by KAON API in purpose of defined a uni-interface ensures compatible data views, with integrated service defined through dynamic instantiation and delegation of underlying external services layer requests relies on the framework provided by Java Management Extension (JMX).

Data and remote Service layer that guarantee access to physical data sets and associates external services such as: reasoning engines and so forth

Text-to-Onto tool suite built initially upon KAON, it helps extracting concepts by providing relevant information for the automatic (and semiautomatic) extraction of ontologies concepts for English, Italian and German languages text documents.

Text-to-Onto is structured on three text mining algorithms: (Sugumaran, 2012)

- Term extraction algorithm.

- Concept association extraction algorithm.

- Ontology pruning algorithm.

A Term extraction algorithm will be used to extract concepts from QAD documents and it provides information such as:

- Frequency.

- Term Frequency Inverse Document Frequency (TFIDF): Shows the importance of a selected term is within a document.

- Entropy: Indicates the rate of disorder of words in a document.

- C-value (Collocation-value): Improves the extraction of nested multi-word terms and collocations.

2.2.2 Translation Software Tool

The research will utilize Google translation as online translator tool in order to translate the document sample from its original language to the destination language, from Arabic to English.

The requirement for translation process is based on the inexistences of software support Arabic languages.

CHAPTER THREE

The research methodology

Overview

This chapter illustrates the proposed research methodology and its related subcomponents in details. In this research, Metadata will be created for QAD of Middle East University (MEU), and comparative experiments for the proposed metadata and the semi-manual one, will be conducted over the same corpus, using the retrieval system used in this quality assurance department.

Introduction

The focus mainly will be on Arabic language documents because it is weak to be utilized by the various automotive ontology and metadata creation tools. Thus organizations that depend on Arabic language documentation would have inefficient document search and retrieval, because of the lack of such tools.

The methodology will utilize an experimental method to address research objectives. A set of comparative experiments will be conducted between two metadata: the semi-manual metadata and the ontology-based metadata using a translation tool (Google Translate). A set of Arabic text documents will be used and applied for the two methods, and human benchmark will judge their outputs. All quantitative results are then will be used to evaluate the process by calculating the error rate and relevant concepts are retrieved, answer the research questions, and prove the hypothesis. Comparing the quantitative results for the two methods will determine their retrieval quality, and determines which one outperformed the other.

The results of doing these experiments have been used to evaluate extraction process and the efficiency of using an automatic translator, where these results have been studied to introduce proper concepts for searching and retrieving the exact information,

and helping to build ontologies. The procedure of the proposed concepts extraction method will persue the following steps and Figure 3-1 shows these steps as a flow chart diagram.

- **Arabic data (text document) collection** from corpuses used in QA department of a university.

- **Building semi-manual metadata** for these Arabic text documents.

- **Document translation** using Automatic translation tool (Google), from Arabic to English language.

- **Ontology concepts extraction** from the English text documents, using KAON environment.

- **Retranslate the extracting concepts**

- **Comparison** over both extracted concepts**.**

- **Results evaluation** for quality (precision) of concepts extraction.

- **Results summarization and representation.**
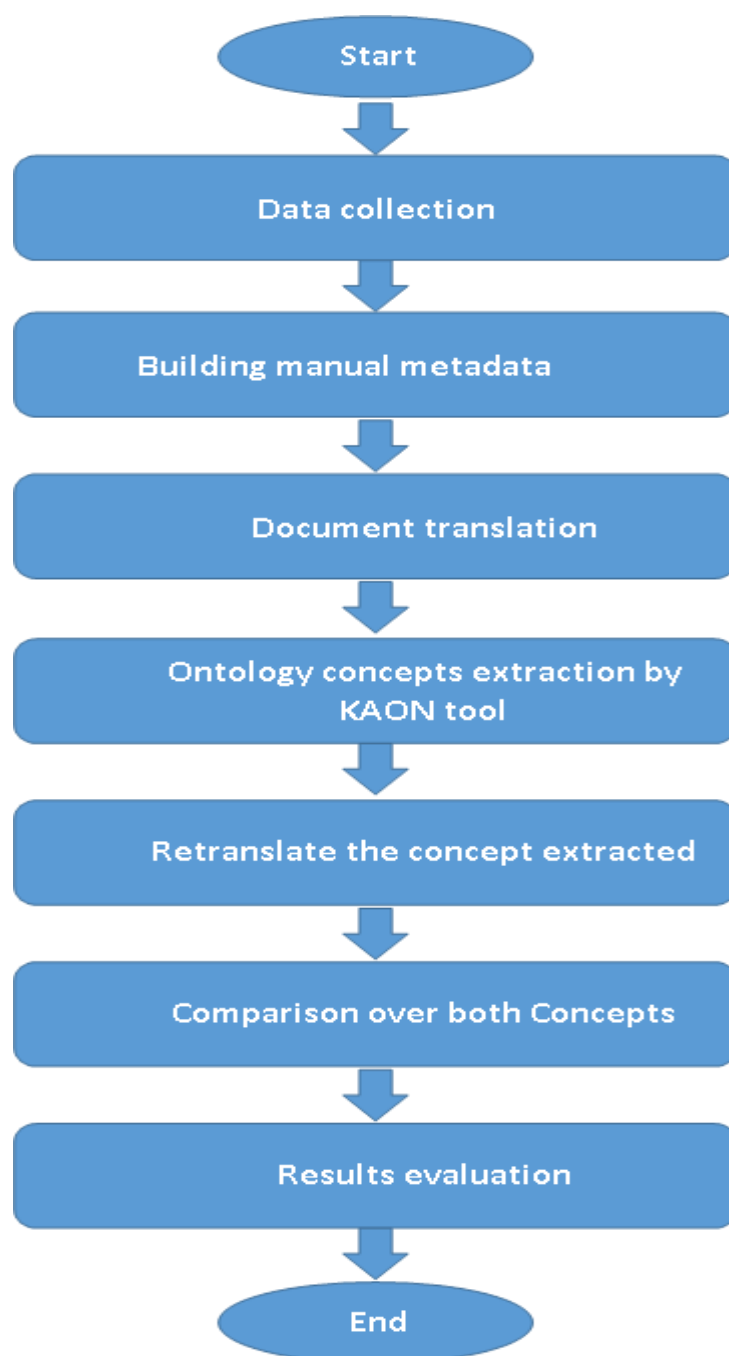
Figure 3-1: Data flow chart for methodology steps.

3.1   Data collection

The primary source of our study is Arabic corpus of Quality Assurance Department (QAD) in the Middle East University (MEU) because QAD considered as the most important sources of information, it contains evidence, research and other valuable materials that can benefit every user.

All Arabic documentation getting from quality assurance department, where the study sample defined in around 200 documents of Adobe PDF and Microsoft Word DOC format.

After examining all Arabic QAD documents, There is difficulty of extracting concept manually from large QA data sets, which consumes a lot of effort and times so the number of documents (sample study) reduced to 20 documents to facilitate the work of this experiment, because of the size of documents, the content of the vocabulary and to ensure the accuracy of the results. Next, the data sample divides the work into three levels to evaluate the results; each level contains a certain number of documents. The three levels start with five documents at level one, then add five documents in level two to be in total ten documents and finally add ten documents to level three to be in total twenty documents to use them in the next phase.

3.2   Building semi-manual metadata

The sample study is an Arabic text documents, and most of its vocabulary related to QAD.

The process of extracting concepts semi-manually is one of the most important stages in our research because of the richness and accuracy of ontology concepts

extraction in this stage will affect all subsequent steps. A semi-manual list of concepts has been created for all three levels:

1. Level one contains 48 Arabic concepts.

2. Level two contains157 Arabic concepts.

3. Level three contains 289 Arabic concepts.

To speed up concepts extraction process, online words counter used to help us select the terms because read and extract concepts manually from documents needs a human's effort, high cost and time. Figure 3-2 shows, the result of online words counter. The way of measuring word frequency is founded on the number of words linked with each document (Lawrence et al., 99). The word frequency prevalence in a paper furnishes a useful measurement of word importance (Luhn, 58). This section performed according to the following steps:

1. Merge the Arabic QAD in each level to a single file.

2. Upload the merge document to the online word counter tool to extract keyword density from the text.

3. Reading the result carefully from high-frequency words to low-frequency words to extract the relevant concepts.

4. Send the result to an expert to evaluate them and return them back.

5. Save the result to use them for the comparison section.

Figure 3-2: The result of online words counter tool

3.3   Document translation

In this section, translate Arabic QAD in each level to the English language by using

a proper available English translation tool to use them in next section. Choosing Google

translate because:

1.   It is free multilingual dictionary machine translation service.

2.   It support file upload.

3.   Easy to use because most people are already familiar with Google Translator.

4.   It is not perfect but still one of the best.

3.4   Ontology concepts extraction

In this section, extracting ontology concepts and their relationship from QAD documents, after translating documents to the English language by Google translate to be supported by our automatic concepts extraction tools and convert all English documents to TXT format file.

KAON Text-To-Onto Tools that consider a suitable tool to help the ontology-building process (Alhasan et al., 2016). The following steps for concepts extraction using this tool:

- Create new corpus function.

- New Term Extraction function.

- Set the frequency threshold parameter.

3.4.1 Create new corpus

Enter the corpus translated text from an experiment sample English text documents to

tool (Text-To-Onto) using new corpus function as shown in figure 3-3.
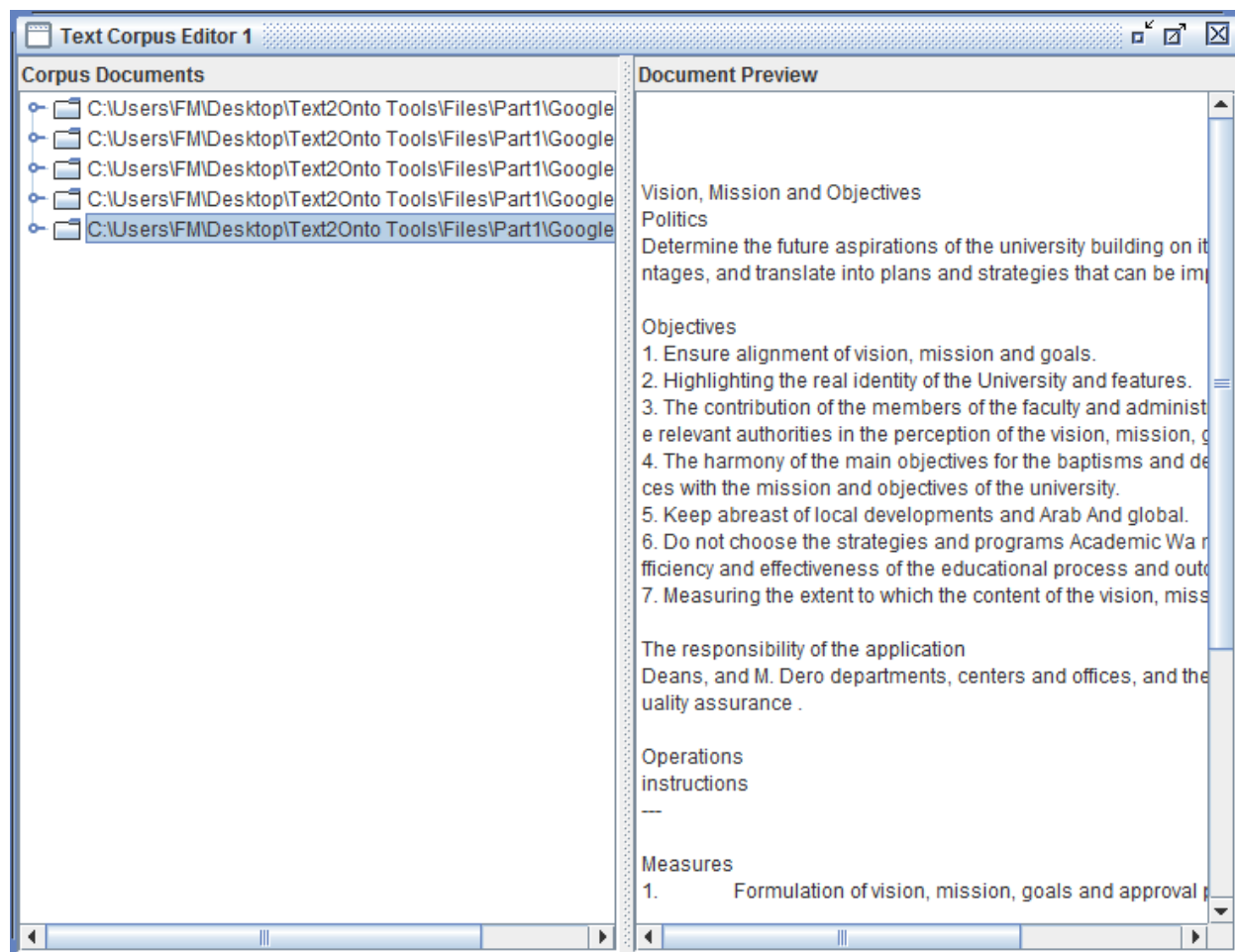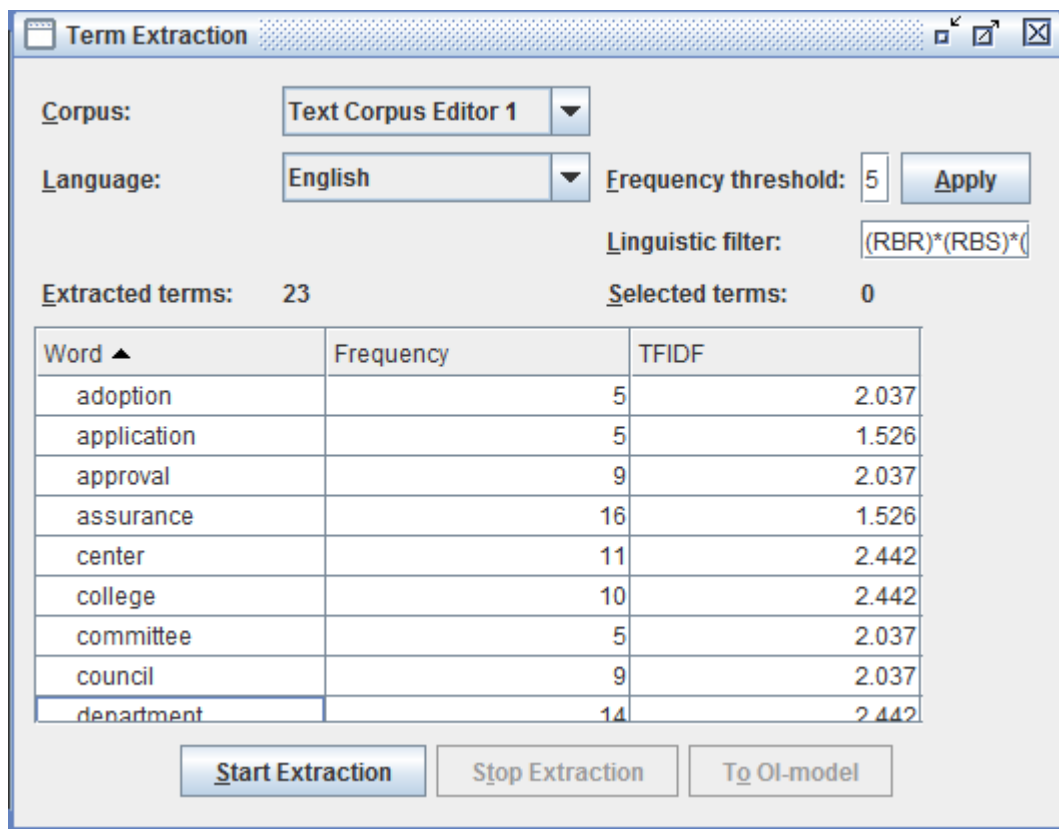
Figure 3-3: New corpus Creation Function Using Text To Onto Tool

3.4.2 New Term Extraction

The Term Extraction module tries to distinguish hopeful terms in the corpus,

which might be utilized as ideas in the scholarly ontology. In this manner for each of

these applicant terms distinctive measures or their recurrence in the corpus are

registered. Then using new term extraction function available in Text To Onto tool to

extract the concept from entered corpus, to extract the concepts in the table as figure 3-4

shown. The extracted word can be sorted based on the various measures such as

Absolute Frequency, TFIDF, Entropy and C-Val, in the research the concepts sorted based on the word Absolute Frequencies as the most suitable measurement for research purpose.

Figure 3-4:Concept Extraction result table.

After extracting the concepts, as obviously seen from the figure 3-4 there is a frequency number corresponding to each concept, based on that we will set a frequency threshold parameter value and count number of concept that has this frequency value.

3.4.3 Set the frequency threshold parameter

In this section, set the frequency threshold parameter to (3) and (5), then check the result respectively. Then keep refined the extraction result through elimination process for multi-words and repeated words (like: quality assurance, college department, department center office) as shown in figure3-5.

Figure 3-5: Elimination process for composite and repeated words.

3.5   Retranslate the concept

In this section, retranslate only the concepts list extracted by KAON tool to Arabic Language using the same automatic translation (Google translate) to ensure compatibility. The result will be stored to be used in the next section.

3.6   Comparison over both Concepts:

In this section, evaluate the efficiency of our method by comparing the output result of concepts extracted by KAON tool and the concepts extracted semi-manually. The evaluation is based on the number of corresponding terms between the concepts extracted by the KAON tool and the concepts extracted semi-manually (Abasca et al., 2003).

In every level, we compare the Arabic concept extracted from Tool, with that extract from same corpus semi-manually. Then extract the tolerance percentage between two methodologies, and compared the tolerance differentiate with related variables such as file size. As shown in figure 3-6.

The comparison decision will be determined in three type:

- Match: The concepts from KAON tool list correspond the concepts of semi-manual list

- Similar: The concept from KAON tool list has the same meaning to one concept of semi-manual list to measure the similarity we depend on dictionary.

- Not matching: The concepts from KAON tool list not found or have different meaning to the concepts of semi-manual list

	Extracting concepts by KAON tool	Translated by Google translator	**Manual** Arabic extracting concepts	
1.	vision	رؤية	الرؤية	Match
2.	mission	مهمة	الرسالة	Similar
3.	quality	جودة	الجوده -النوعية	Match
4.	office	مكتب، مقر، مركز	مكتب	Match
5.	assurance	توكيد	ضمان	Similar
6.	department	قسم، أقسام	الأقسام	Match
7.	center	مركز	المركز	Match
8.	college	كلية	الكلية	Match
9.	council	مجلس	مجلس	Match
10	approval	موافقة	الموافقة	Match
11	university	جامعة	الجامعة	Match
12	model	نموذج	نموذج	Match
13	director	مدير	مدير	Match
14	adoption	تبنى	إقرار	**Not Match**

Figure 3-6: Comparison over both metadata.

3.7 Results evaluation

The results of concept extraction tool are evaluated by examining the results of the tool by the relevance judgments of a human expert. This means that we have to classify the outcome of the tools in relevant or irrelevant concepts. This idea leads to use the evaluation scheme based on information retrieval:

$$Error\ rate = \frac{a}{b} \tag{1}$$

Where "a" is number of wrong concepts extracted by the KAON tool (NOT MATCH), "b" is the total number of concepts extracted by the KAON tool.

And to measure the percentage of only the relevant concepts we used:

$$Precision = \frac{c}{c+d} * 100\% \tag{2}$$

Where "c" is the number of concepts extracted semi-manually, "d" is number of wrong concepts extracted by the KAON tool. Precision is the percentage of only the relevant concepts, even though it skips irrelevant concepts. Therefore, higher Precision indicates that most of the relevant concepts are retrieved. (Abasca et al., 2003)

By evolution the last Precision Formula through two case, first that there is no wrong concepts extracted which mean the precision must be 100% logically, through numerical calculated via formula by assuming d= 0 as bellow:

$$Precision\ (no\ wrong\ concpts\ extracted) = \frac{c}{c+0} * 100\% = 100\% \qquad (3)$$

In this case, the numerical finding indicates the logical expectation.

The second case is assuming that all concepts extracted were wrong thus, the Precision must be zero as logically expectations. While going through numerical calculation by assuming d = c the as bellow

$$Precision\ (all\ concpts\ extracted\ were\ wrong) = \frac{c}{c+d} * 100\%$$

$$= \frac{1}{2} * 100\% = 50\% \qquad (4)$$

Based on the formula the precision is 50%, which is wrong logically, thus, this formula was not sufficient for all cases. After continuous enhancement procedure made, the precision formula became as bellow:

$$Precision = \frac{c}{c + d + |E|} * 100\% \qquad (5)$$

Where the "c" is the number of concepts extracted semi-manually, "d" is number of wrong concepts extracted by the KAON tool, and "E" is the difference between the semi-manual concepts number and number of concepts extracted by the KAON tool, where E can be calculated as bellow:

$$|E| = |C - the\ number\ of\ all\ Kaon\ extracted\ concepts| \qquad (6)$$

CHAPTER FOUR

Experimental Results & Evaluation

Overview

This chapter discusses in details, the results of a set of comparative experiments between the semi-manual metadata and the ontology-based metadata by using a set of Arabic text documents for the two methods, where the Arabic document data is converted to English language using an automatic translation tool (Google translate).

At the beginning the concepts will be extracted semi-manually from the original corpus, after collecting the concepts and arranged in tables, the second phase will be extracting concepts from translated corpus automatically, the documents sample distributed over three levels as bellow:

- Level One: consist of five documents.

- Level Two: consist of five new documents plus previous 5 documents.

- Level Three: consist of ten documents plus previous 10 documents.

4.1 Semi-Manual concepts extraction:

The concept extracted semi-manually for all prepared texts corpus by using tool that extract keyword density and frequency of the words to count the number of words that are repeated in all documents. The extracting concepts will be utilized for comparison purpose with automatic extracted concept in all the three levels.

4.2   Automatic concepts extraction

The Automatic extraction utilized KAON Text To Onto tool software that is considered a suitable tool in order to support ontology building process, KAON Text To Ontology tool does not support Arabic language, thus the need was emerged to translate the original corpus to other supported language tool such as, English. The research used Google translate online tool for translation purpose. Appendices includes sample of the original Arabic texts and its corresponding translated text. Furthermore, the Text-To-Onto tool support format (.TXT), where the sample documents format is (.DOCX), thus all documents are converted to the (.TXT) format.

4.3   Comparison over both concepts extraction

The execution of the KAON tool measured through comparing it with output concepts extracted semi-manually .The measure of execution depends on the quantity of relating terms between the concepts extracted by the tool and the concepts extracted semi-manually

4.4   Level one:

It is the first stage of extraction process characterized as:

1. Experiment Sample: five Documents.

2. Documents Size : 6.61 KB

3. Total Words: 967

4.4.1 Semi-Manual Extraction:

The results of extracting concepts semi-manually reach 48 terms, shown in table 4-1.

Table 4-1: The extracted terms semi-manual for five documents

اجتماعات	إجراءات	إقرار-إقرارها	أداء
أعضاء	أقسام	أهداف	تُعرض
تبني	تحقق	تدريسية	تطبيق
تطويرها	تعديل-تعديلات	تعرض	تقارير
تنفيذ	توصيات	جامعة	جهات
جودة – نوعية	دراسة-دراسات	رسالة	رؤية
صيغة	ضمان	طالب	علاقة
عمداء	عمل	عمليات	قسم-دائرة
كفاءة	كلية-كليات	لجنة	متعلقة
مجلس	هدف	محاضر	مدير
مركز-مراكز	مسؤولية	مكتب-مكاتب- مركز	موافقة
مؤشر-مؤشرات	نسخة	نموذج	نهائية

4.4.2 Frequency threshold parameter (5) :

4.4.2.1   Automatic Extraction:

The result of extracting concepts from documents by KAON tool is 30 terms, shown in table 4-2:

Table 4-2:Concept extracting by KAON tool threshold (5)

adoption	application	approval	assurance
authorities	center	college	committee
council	deans	department	director
directors	goal	indication	indicator
meetings	mission	model	objectives
office	operations	procedures	quality
responsibility	sections	university	version
goals	vision		

4.4.2.2   Extraction Comparison:

The comparison between two methods shown in table 4-3 below:

Table 4-3:Comparison over both concepts extraction list threshold (5)

	English concepts extraction by KAON tool	Translated concepts to Arabic by Google translate	Arabic concepts extraction semi-manually	Matching
1.	adoption	تبني	تبني	Match
2.	application	الوضعية	التطبيق	Not Match
3.	approval	موافقة	الموافقة	Match
4.	assurance	توكيد	ضمان	Similar
5.	authorities	السلطات	الجهات	Similar
6.	center	مركز	المركز	Match
7.	college	كلية	الكلية	Match
8.	committee	لجنة	لجنة	Match

9.	council	مجلس	مجلس	Match
10	deans	عمداء	العمداء	Match
11	department	قسم، أقسام	القسم-الدائرة	Match
12	director	مدير	مدير	Match
13	directors	الإدارة		Not Match
14	goal	هدف	الهدف	Match
15	goals	أهداف	أهداف	Match
16	indication	إشارة		Not Match
17	indicator	مؤشر	المؤشر	Match
18	meetings	اجتماعات	اجتماعات	Match
19	mission	مهمة	رسالة	Similar
20	model	نموذج	نموذج	Match
21	objectives	الأهداف	الأهداف	Match
22	office	مكتب. مقر. مركز	مكتب - مركز	Match
23	operations	عمليات	العمليات	Match
24	procedures	الإجراءات	الإجراءات	Match
25	quality	جودة	الجودة - النوعية	Match
26	responsibility	المسئولية	مسؤولية	Match
27	sections	الأقسام	الأقسام	Match
28	university	جامعة	الجامعة	Match
29	version	الإصدار		Not Match
30	vision	رؤية	الرؤية	Match

Automatic extraction count 30 terms for threshold (5) ,   The composition between semi-manual concepts extraction (48 terms) and KAON tool shows 23 *matched* terms, 3 *similar* terms and 4 *not matched* terms from 30 terms extracted by KAON tool that means that the error rate= (4 / 30) ≈ 0.13 and Precision= (48/(48+4+18)) ≈ 0.69.

4.4.3 Frequency threshold parameter (3) :

4.4.3.1   Automatic Extraction:

 The result of extracting concepts from documents by KAON tool is 50 terms, shown in table 4-4 :

Table 4-4:Concept extracting by KAON tool threshold (3)

action	adjustments	adoption	amendments
application	approval	assurance	authorities
board	center	college	committee
copy	council	deans	department
development	director	directors	display
efficiency	faculty	formulation	goal
implementation	indication	indicator	meetings
members	mission	model	objectives
office	operations	performance	procedures
quality	recommendations	reports	responsibility
sections	students	studies	study
target	teams	transfer	university
version	vision		

4.4.3.2 Extraction Comparison:

The comparison between two methods as shown in table 4-5 below:

Table 4-5:Comparison over both concepts extraction list threshold (3)

	English concepts extraction by KAON tool	Translated concepts to Arabic by Google translate	Arabic concepts extraction semi-manually	Matching
1.	action	عمل	عمل	Match
2.	adjustments	التعديلات	تعديلات	Match
3.	adoption	تبني	تبني	Match
4.	amendments	تعديلات	تعديلات	Match
5.	application	الوضعية	التطبيق	Not Match
6.	approval	موافقة	الموافقة	Match
7.	assurance	توكيد	ضمان	Similar
8.	authorities	السلطات	الجهات	Similar
9.	board	مجلس	مجلس	Match
10	center	مركز	المركز	Match
11	college	كلية	الكلية	Match
12	committee	لجنة	لجنة	Match
13	copy	نسخ	نسخة	Match
14	council	مجلس	مجلس	Match
15	deans	عمداء	العمداء	Match
16	department	أقسام قسم،	القسم-الدائرة	Match
17	development	تطوير	تطويرها	Match
18	director	مدير	مدير	Match
19	directors	الإدارة		Not Match

20	display	عرض	تعرض	Match
21	efficiency	نجاعة	الكفاءة	Not Match
22	faculty	كلية	الكلية	Match
23	formulation	صيغة	الصيغة	Match
24	goal	هدف	الهدف	Match
25	implementation	التنفيذ	تنفيذ	Match
26	indication	إشارة		Not Match
27	indicator	مؤشر	المؤشر	Match
28	meetings	اجتماعات	اجتماعات	Match
29	members	أفراد	أعضاء	Match
30	mission	مهمة	رسالة	Similar
31	model	نموذج	نموذج	Match
32	objectives	الأهداف	الأهداف	Match
33	office	مكتب. مقر. مركز	مكتب - مركز	Match
34	operations	عمليات	العمليات	Match
35	performance	أداء	الأداء	Match
36	procedures	الإجراءات	الإجراءات	Match
37	quality	جودة	الجودة - النوعية	Match
38	recommendations	توصيات	**التوصيات**	Match
39	reports	تقارير	تقارير	Match
40	responsibility	المسئولية	مسؤولية	Match
41	sections	الأقسام	الأقسام	Match
42	students	الطلاب	طالب	Similar
43	studies	دراسات	الدراسات	Match
44	study	دراسة	الدراسة	Match
45	target	استهداف	الهدف	Match

46	teams	فرق		Not Match
47	transfer	تحويل		Not Match
48	university	جامعة	الجامعة	Match
49	version	الإصدار		Not Match
50	Vision	رؤية	الرؤية	Match

The terms extracting are 50 concepts from KAON tool. The composition between semi-manual concepts extraction (48 terms) and KAON tool shows 39 *matched* terms, 4 *similar* terms and 7 *not matched* terms from 50 terms extracted by KAON tool that means that the error rate= (7 / 50) ≈ 0.14 and Precision= (48/(48+7+2)) ≈ 0.84.

The decision making: the automatic translation with threshold (3) has minimum error rate equal to 14% and Precision equal to 84% where the error rate for automatic translation with threshold (5) has an error rate 13% and Precision equal to 69%. As shown in figure 4-1.

Figure 4-1: Level (1) Error rate and Precision for frequency threshold (3) and (5)

4.5 Level Two:

It is the second stage of extraction process characterized as**:**

1. Experiment Sample: Ten Documents.

2. Documents Size : 32.5 KB

3. Total Words: 2811

4.5.1 Semi-Manual Extraction:

Part of the results of extracting concepts semi-manually reach 157 terms, shown in table

4-6. For more details see table in the appendix:

Table4-6: Semi-Manual Extracted Terms

اجتماعات	استجابة	استعمال	اسم
اعتماد	امتحان	إجابة	إجراء - إجراءات
إداري	إعادة	إقرار-إقرارها	إنجليزية
أداء	أردني	أستاذ	أعضاء
أقسام	أكاديمي	أمناء	أهمية
أوافق	باحث - باحثون	بحث -بحوث	بند
بيانات	تُعرض	تُلغى	تاريخ
تبني	تحقق	تخصص	تدريسية
ترفق	ترقية	تسجيل	تطبيق
تطويرها	تعديل-تعديلات	تعرض	تعليمات
تعيين	تقرير	تقوم	تقييم
تلغى	تنفيذ	توصية-توصيات	توقيع
جامعة	جامعي	جلسة	جهات
جودة - نوعية	خارجي	خطة	دراسات-دراسة
درجة	دعم	دقيق	دينار

4.5.2 Frequency threshold parameter (5) :

4.5.2.1 **Automatic Extraction**:

Part of the result of extracting concepts from documents by KAON tool is 101 terms, shown in table 4-7. For more details see table in the appendix:

Table 4-7: Concept extracting by KAON tool for threshold (5)

academic	acceptable	accepted	adjustment
adjustments	admission	admissions	adoption
amended	amendments	answer	application
approval	Arabic	assistant	assurance
author	authorities	board	book
center	clarity	college	committee
contents	copy	council	date
dean	deanship	decision	department
dinars	director	directors	discussion
estimated	evaluation	exam	faculty
fee	final	form	forms
goal	graduate	grant	history

4.5.2.2 Extraction Comparison:

Part of the comparison between two methods shown in table 4-8. For more details see table in the appendix.

Table 4-8:Comparison over both concepts extraction list threshold (5)

	English concepts extraction by KAON tool	Translated concepts to Arabic by Google translate	Arabic concepts extraction semi-manually	Matching
1.	academic	أكاديمي	أكاديمي	Match
2.	acceptable	مقبول	مقبول	Match
3.	accepted	قبلت	مقبول	Match
4.	adjustment	تعديل	تعديل	Match
5.	adjustments	التعديلات	تعديلات	Match
6.	admission	قبول	القبول	Match
7.	admissions	القبول	القبول	Match
8.	adoption	تبني	تبني	Match
9.	amended	معدل	معدل	Match
10.	amendments	تعديلات	تعديلات	Match

The terms extracted concepts from KAON tool are 101. The composition between semi-manual concepts extraction (157 terms) and KAON tool shows 84 *matched* terms, *7 similar* terms and *10 not matched* terms from 101 terms extracted by KAON tool that means that the error rate=(10/101)≈0.10 and Precision= (157/(157+10+56)) ≈ 0.70.

4.5.3 Frequency threshold parameter (3) :

4.5.3.1 **Automatic Extraction**:

Part of the result of extracting concepts from documents by KAON tool is 170 terms, shown in table 4-9. For more details see table in the appendix:

Table 4-9:Concept extracting by KAON tool threshold (3)

academic	acceptable	accepted	action
address	adjustment	adjustments	administrative
admission	admissions	adoption	amended
amendments	amount	answer	Arabic
application	appropriate	approval	assurance
arbitrator	author	assistant	board
attached	canceled	authorities	chapter
book	college	center	conditions
clarity	contribution	committee	council
contents	date	copy	dean

4.5.3.2 **Extraction Comparison**

Part of the comparison between two methods as shown in table 4-10. For more details see table in the appendix:

Table 4-10:Comparison over both concepts extraction list threshold (3)

	English concepts extraction by KAON tool	Translated concepts to Arabic by Google translate	Arabic concepts extraction semi-manually	Matching
1.	academic	أكاديمي	أكاديمي	Match
2.	acceptable	مقبول	مقبول	Match
3.	accepted	قبلت	مقبول	Match
4.	action	عمل	عمل	Match
5.	address	عنوان	عنوان	Match
6.	adjustment	تعديل	تعديل	Match
7.	adjustments	التعديلات	تعديلات	Match
8.	administrative	إداري	الإداري	Match
9.	admission	قبول	القبول	Match
10.	admissions	القبول	القبول	Match

The terms extracted from KAON tool are 170. The composition between semi-manual concepts extraction (157 terms) and KAON tool shows 136 *matched* terms, 13 *similar* terms and 21 *not matched* terms from 170 terms extracted by KAON tool that means that the error rate= (21 / 170) ≈ 0.12 and Precision= (157/(157+21+13)) ≈ 0.82.

The decision making: The automatic translation with threshold (3) has minimum error rate equal to 12% and Precision equal to 82% where the error rate for automatic translation with threshold (5) has an error rate 10% and Precision equal to 70%, As shown in figure 4-2.
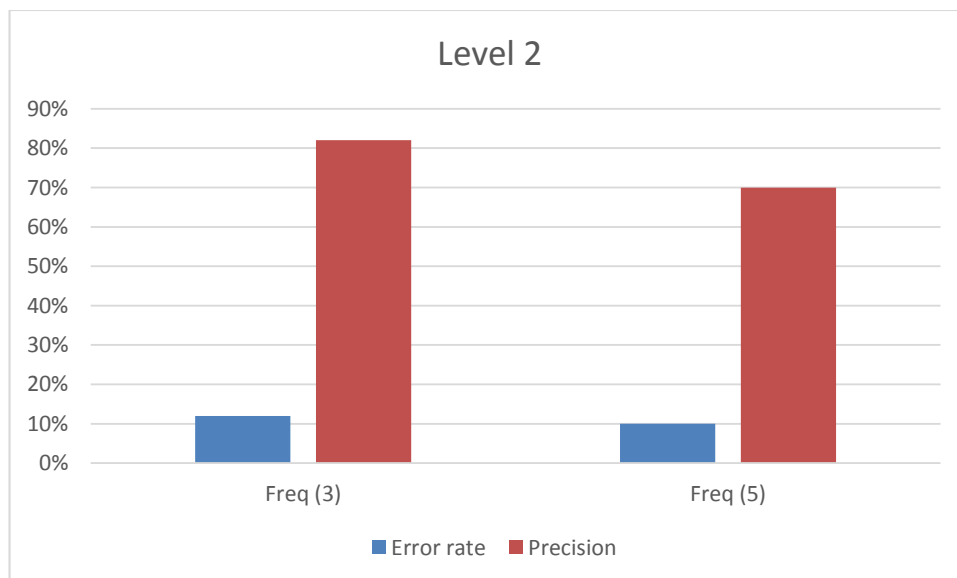
Figure 4-2: Level (2) Error rate and Precision for frequency threshold (3) and (5)

4.6   Level Three:

It is the third stage of extraction process characterized as**:**

1. Experiment sample: 20 Documents.

2. Documents Size : 68.0 KB

3. Total words : 5796

4.6.1 Semi-Manual Extraction:

 Part of the result count 289 terms as shown in table 4-11. For more details, see table in the appendix:

Table 4-11:Semi-Manual Extraction Concept For

اجتماعات	اجتهاد	اختبارات	اختياري
استبانة	استجابة	استطلاع	استعمال
اسم	اعتماد	اقتراحات	اقرار
الإلكتروني	التحاق	التزام	التسديد
الطالب	اللازمة	اللغة	المالي - المالية
المعنية	امتحان	انترنت	انذار
انسحاب	اهتمام	إجابة	إجراء - إجراءات
إخطار	إدارة	إداري	إشارة
إشعار	إعادة	إعداد	إقرار
إنتاجية	إنجليزية	إنذار	أتعهد
أجهزة	أخرى	أخلاق	أداء
أردني	أسباب	أستاذ	أعضاء
أقسام	أكاديمي	أمناء	أنظمة

4.6.2 Frequency threshold parameter (5) :

4.6.2.1   Automatic Extraction:

Part of the result of extracting concepts from documents by KAON tool is 174 terms, shown in table 4-12. For more details see table in the appendix:

Table 4-12:Concept extracting by KAON tool threshold (5)

Absence	academic	accepted	access
achieved	action	activity	adjustment-adjustments
administrative	admission-admissions	adoption	amended
Amount	answer	application	applications
approval	Arabic	assessed-assessment	assurance
Author	authorities	board	book
Center	chapter	clarity	coach
College	commitment	committee	compatibility

4.6.2.2 Extraction Comparison:

Part of the comparison between two methods is shown in table 4-13. For more details see table in the appendix:.

Table 4-13:Comparison over both concepts extraction list threshold (5)

	English concepts extraction by KAON tool	Translated concepts to Arabic by Google translate	Arabic concepts extraction semi-manually	Matching
1.	absence	غياب		Not Match
2.	academic	أكاديمي	أكاديمي	Match
3.	accepted	قبلت	مقبول	Match
4.	Access	التمكن من	الدخول	Match
5.	achieved	حقق	تحقق	Match
6.	Action	عمل	عمل	Match
7.	Activity	نشاط	نشاط	Match
8.	adjustment-adjustments	تعديل	تعديل-التعديلات	Match
9.	administrative	إداري	**الإداري**	Match
10.	admission-admissions	قبول	القبول	Match

The terms extracting from KAON tool are 174 terms, The composition between semi-manual concepts extraction (289 terms) and KAON tool shows 145 *matched* terms, *18 similar* terms and 11 *not matched* terms from 174 terms extracted by KAON tool that means that the error rate= (11 / 174) ≈ 0.06 and Precision= (289/(289+11+115)) ≈ 0.70.

4.6.3 Frequency threshold parameter (3) :

4.6.3.1 Automatic Extraction:

Part of the result of extracting concepts from documents by KAON tool is 323 terms, shown in table 4-14. For more details see table in the appendix:

Table 4-14: Concept extracting by KAON tool threshold (3)

absence	academic	acceptable	accepted
access	account	achieved	action
activity	address	adequate	adjustment-adjustments
administrative	admission-admissions	adoption	affairs
allocated	amended	amount	answer
application	applications	appropriate	approval
approving	Arabic	arbitrator	assessed-assessment
assurance	attached	attention	audit
author	authorities	availability	average

4.6.3.2 Extraction Comparison

Part of the comparison between two methods is shown in table 4-15. For more details, see table in the appendix:

Table 4-15: Comparison over both concepts extraction list threshold (3)

	English concepts extraction by KAON tool	Translated concepts to Arabic by Google translate	Arabic concepts extraction semi-manually	Matching
1.	absence	غياب		Not Match
2.	academic	أكاديمي	أكاديمي	Match
3.	acceptable	مقبول	مقبول	Match
4.	accepted	قبلت	مقبول	Similar
5.	access	التمكن من	دخول	Match
6.	account	الحساب	حساب	Match
7.	achieved	حقق	تحقق	Match
8.	action	عمل	عمل	Match
9.	activity	نشاط	نشاط	Match
10.	address	عنوان	عنوان	Match

The extracted terms from KAON tool are 323 terms, The composition between semi-manual concepts extraction (289 terms) and KAON tool shows 230 *matched* terms, 48 *similar* terms and 45 *not matched* terms from 323 terms extracted by KAON tool that means that the error rate= (45 / 323) ≈ 0.14 and Precision= (289/(289+45+34)) ≈ 0.79.

The decision making : The automatic translation with threshold 3 has minimum error rate equal to 14% and Precision equal to 79% where the error rate for automatic translation with threshold 5 has an error rate 6% and Precision equal to 70%, As shown in figure 4-3.
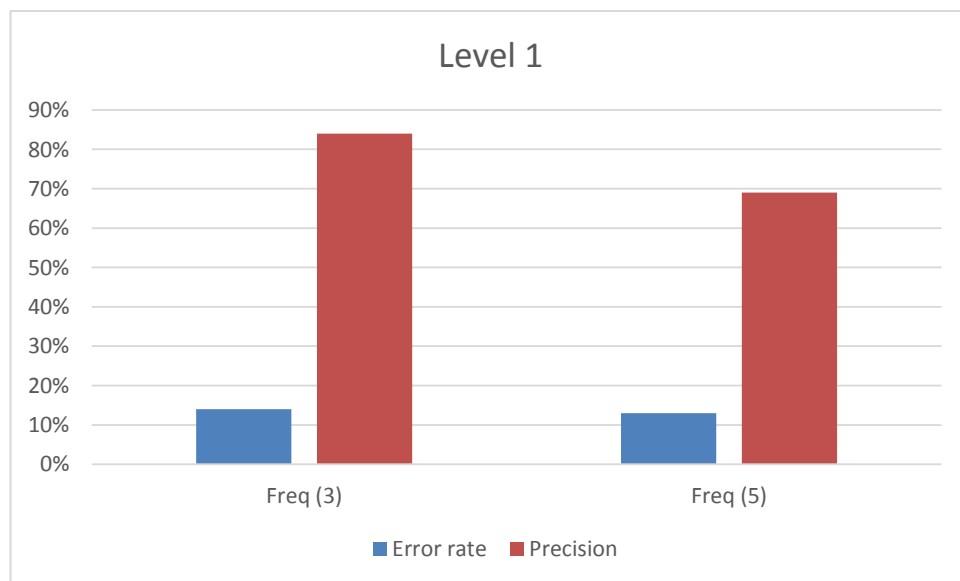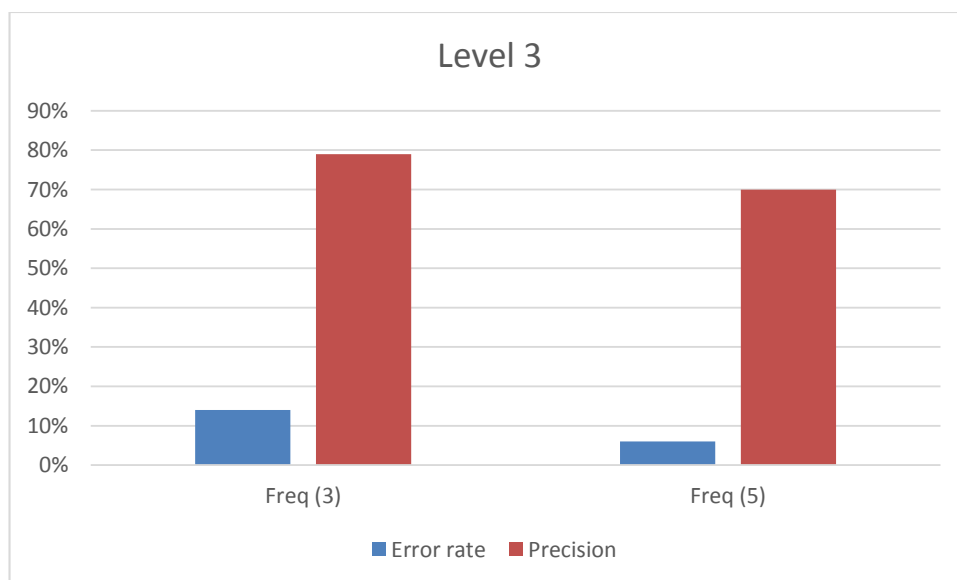
Figure 4-3: Level (3) Error rate and Precision for frequency threshold (3) and (5)

CHAPTER FIVE

Conclusions and Future Work

Overview

This chapter discusses the result of experimental procedures done through this research, and it highlights some recommendations.

5.1 Conclusion

Through all three levels of experiments, the error rate for same frequency threshold such as (3) performed error rate 14%, 12%, 14% and precision rate 84%, 82%, 79% respectively for documents number 5,10,20 per level, and on threshold such as (5) performed error rate 13%, 10%, 6% respectively and precision rate 69%, 70%, 70%, that show constant behaviors based on documents sample. As shown in figure 5-1.
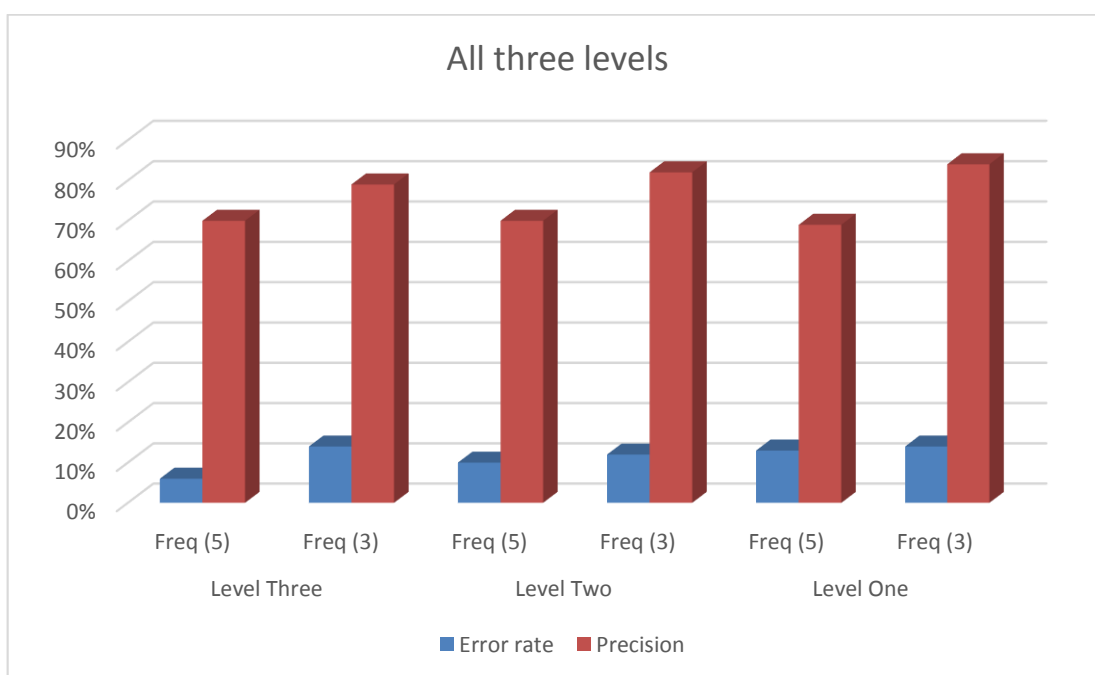
Figure 5-1: All Levels Error rate and Precision for frequency threshold (3) and (5)

As research shows the error rate and precision rate are depends on the frequency thresholds, where the numerical results show high value of the precision rate, also the error rate decreases gradually as increasing the document sampling.

Based on the comparison process, the Frequency thresholds results in maximum error rate condition registered 86% success rate, and 94% success rate in the advanced conditions (minimum error rate). In addition, the minimum number of relevant concepts are retrieved 69%; and the maximum number of relevant concepts are retrieved 84%.

However, the result shows an ambiguous relation and sensitivity between the Precision and the document sample size, where it is not clear to identify the related behaviors between them, the research suggested further researches and attempts to state the relation between them which surely effect on the validity of suggested approach for huge documents.

The experiments prove that the proposed technique achieved the highest success rate and enhancement results associate to usage the proposed frequency Threshold on (5).

The previously mentioned percentages considered high and valuable percentages in the area of ontology based Arabic concept that obtained from unsupported Arabic language ontology software. Which means using automatic translation tools (Google translate) and KAON ontology building for extracting concept in the QA domain is considered a suitable and power tool enhanced retrieving and supported ontology-based metadata build basis for Arabic documents compared to the semi-manual method.

5.2   Recommendation

Through the conduct of the research's idea, many issues and challenges were raised such as time, resources and other limitations. The research suggests a few recommendations for the ability to use the proposed approach in extracting concepts to help build a full ontology for QA domain. Promoting the KAON ontology software utilization for extracting concept to help build ontologies in QA domain and using automatic translation tool (Google translate) to translate the Arabic documents to support the extraction tool, where this approach promotes and enhances the quality assurance system potentials.

The research experimental results proved the efficiency of Google translate and KAON Tool used for extraction purpose and the highest correlation value compared with semi-manual methods, lowest error rate, and less time and effort operations especially for huge database.

Furthermore, the research shows the differences in results for using KAON ontology building tool for extracting concepts with various frequencies related to QA domain.

5.3   Contribution

Despite the large and widespread of ontologies building software and its frequently studied and mentioned in researches, the raised issue that until now there is no official integrated ontology software supports the Arabic Language, thus some methodology were suggested and examined to cope this issue through exploit the translation tools in view of optimum utilization of the current ontology building software.

The novel of the methodology proposed the basis of extracting the Arabic concepts process for ontology building software purpose that concentrates in promoting a new approach supporting Arabic Language via Translation Tool, enhancing the efficiency of existence software utilization, and extends the extracting concepts vision.

The research results contribute to demonstrate that Translation tools are good alternative approach come over the weakness of language-supported issues in the ontology-building domain, thus the research attend to make contribution to the experimental literature on goals, methodology and variety.

5.4   Future research

As research showed, the extracting concepts in Arabic corpus were affected by performance of translation tools and similarity measures. Thus, the research suggest that:

- Possibility to reform the experiment using various Translation tools

- Build full ontology for Quality Assurance based on concept extracted

- Possibility of reforming the experimental considers the similarity measurement such as TFIDF, Entropy and C-Val.

- Refine the frequency threshold for extracting concepts.

- Develop the QA system for exploring documents.

- Deployed the research approaches on various domains.

- Develop ontology-based metadata to support or enhance retrieving in Arabic QA documents compared to the semi-manual metadata.

This research will extend the vision of extracting concept for building ontology purposes especially for Arabic corpus context, which consider a big challenges faced the Arabic building ontologies, and provide a promising solution in extracting and building ontology for huge systems.

5.5   Project schedule and device specifications:

The figure 5-2 shown the working schedule.
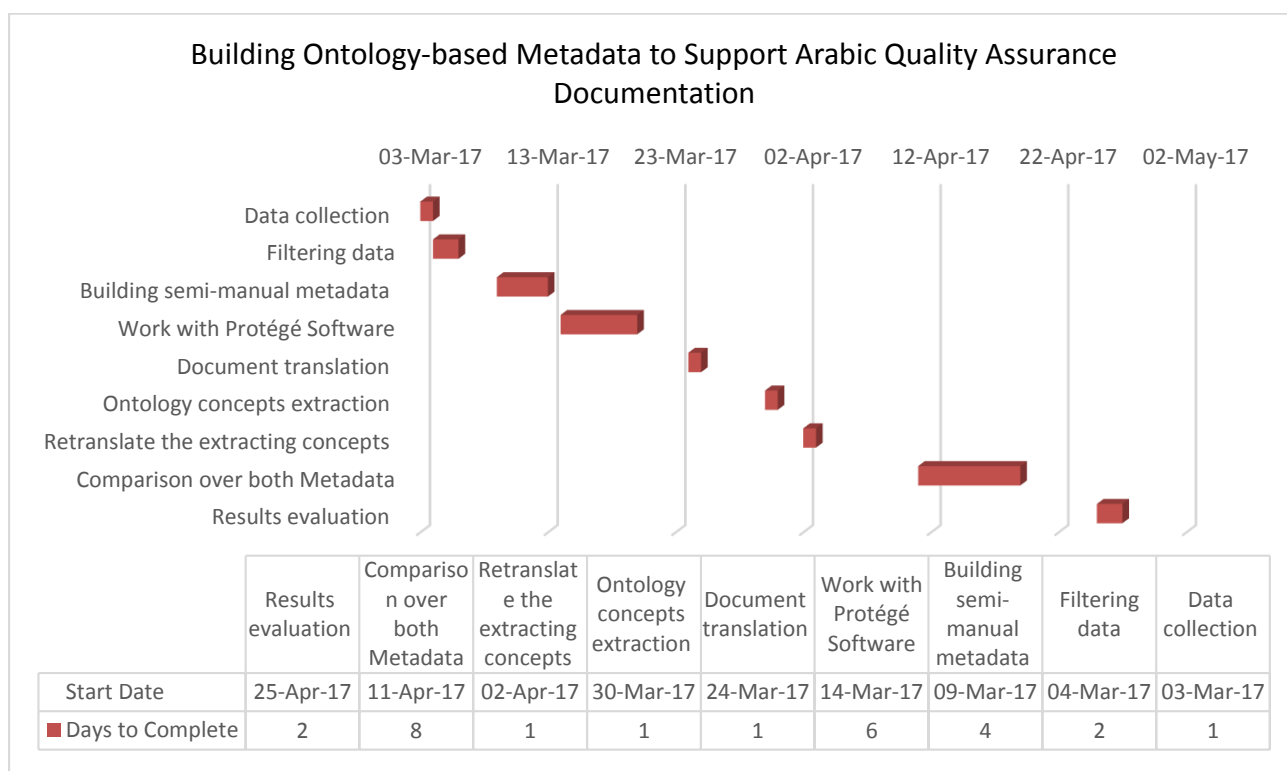
Figure 5-2: Project work schedule

Machine specification used in the study:

- System Manufacturer: LENOVO

- System Model: 80G0

- BIOS: InsydeH2O Version 05.03.40A7CN47WW

- Processor: Intel(R) Celeron(R) CPU N2840 @ 2.16GHz (2 CPUs), ~2.2GHz

- Memory: 2048MB RAM

- Operating System: Windows 10 Enterprise 2016 LTSB 32

References

References

- Abascal-Mena, R., & Rumpler, B. (2007). Semantic content annotation and ontology creation to improve pertinent access to digital documents. In *4th Italian Semantic Web Workshop, Bari, Italy*.

- Abascal, R., Rumpler, B. and Pinon, J. (2003). "*An Analysis of Tools for an Automatic Extraction of Concept in Documents for a Better Knowledge Management*". France: Villeurbanne cedex.

- Ahmed, S. (2014). *Extracting instances to Improve semantic matching in the domain of Renewable Energy*, (Unpublished Master thesis), Middle East University, Amman, Jordan.

- Aktas, M., Oh, S., Fox, G., & Pierce, M. (2006). "XML Metadata Services. Semantics, Knowledge". *Second International Conference on Semantics, Knowledge and Grid*, USA:IEEE.

- Aldiery, M. (2017). *The Semantic Similarity Measures Using Arabic Ontology*, (Unpublished Master thesis), MIDDLE EAST UNIVERSITY , Amman, Jordan.

- Alhasan, M. Noaman, & AL-Hassan, M. (2016). Towards an Ontology of Data Breaches Threat for Cloud Computing, *International Journal of Advanced Research in Computer and Communication Engineering IJARCCE*, 5 (5).

- Alkalifa, M. & Rodriguez, H. (2010).Automatically Extending Named Entities Coverage of Arabic WordNet using Wikipedia, *International Journal on Information and Communication Technologies*, 3(3).

- Araki, K., Ohashi, K., Yamazaki, S., Hirose, Y., Yamashita, Y., Yamamoto, R., Minagawa, K., Sakamoto, N. & Yoshihara, H. (2000). Medical Markup Language (MML) for XML-based Hospital Information Interchange. *Journal of Medical Systems,* 24(3), pp. 195-211.

- Benson, C. (2012, March 29). *An Intro to Metadata and Taxonomies*. (on line) Available: /http://blog.braintraffic.com/2012/03/an-intro-to-metadata-and-taxonomies/

- Bozak, E., Ehrig, M., Handschuh, S., Hotho, A., Maedche, A., Motik, B., Oberle, D., Schmitz, C, Studer, R., Stumme, G., Sure, Y., Staab, S., Stojanovic, L., Stojanovic, N., Tane, J., Volz, R.& Zacharias, V. (2002). KAON - Towards a large scale Semantic Web, *E-Commerce and Web Technologie.* Springer, pp. 231-248.

- Bui, H. (2012). *A rich metadata file system for scientific data*. (unpublished dissertation of doctor philosophy), University of Norte Dame, Indiana, lll: USA.

- Cjien-cheng, L. & Chao-chen, C. (2009). Archiving and Management of Digital Images Based on an Embedded Metadata Framework. *Proc. Int'l Conf. on Dublin Core and Metadata Applications*, pp. 71-84.

- Gabel, T. Sure, Y. & Voelker, J. (2004). KAON – An Overview Karlsruhe Ontology Management Infrastructure, FZI – *Research Center for Information Technologies*, Karlsruhe at Institute AIFB, University of Karlsruhe.

- Hokstad, T .(2015). *Ontology based study planning and classification of university subjects*, unpublished thesis, university of Agder.

- IPCC (2000). "Good practice Guidance and uncertainty Management in National Green House Gas Inventors". *In: The Intergovernmental Panel on Climate Change*, Switzerland, pp 8-11.

- Johnston, P. (2005). *"Metadata Sharing and XML*". nof-digitise Technical Advisory Service, (online), available: http://www.ukoln.ac.uk/nof/support/help/papers/metaxml/

- Jurˇc´ık ﹐ A. (2010). Development of Visualization Plug-in for Protégé, unpublished master thesis**,** *MASARYKOVA UNIVERZITA FAKULTA INFORMATIKY.*

- Kontokostas, D., Mader, C., Dirschl, C., Eck, K., Leuthold, M., Lehmann, J. & Hellmann, S. (2016). Semantically Enhanced Quality Assurance in the JURION Business Use Case. *13ᵗʰ international conference*, ESWC 2016,Herakilio, Greece , May 29- June 2, 2016 , PP 661-676, Switzerland :Spring international publishing, Vol. 9678, DOI 10.1007/978-3-319-34129-3_40.

- Lawrence S., Lee Giles C., & Bollacker K. (1999). Digital Libraries and Autonomous Citation Indexing, *IEEE Computer*, 32(6), pp 67-71.

- Lee, P. (2003). *Metadata Representation and Management for Context Mediation*. (unpublished Master thesis), Massachusetts Institute of Technology. Massachusetts, UK

- Luhn, H.P. (1958). The automatic creation of literature abstracts, *IBM Journal of Research and Development*, 2, pp.159- 165.

- Manghani, K. (2011). Quality assurance: Importance of Systems and standard operating procedures, *Perspectives in Clinical Research*, 2(1): pp 34-37.
- Mckemmish, S. & Acland, G. (1999). Accessing Essential Evidence On The Web: Towards An Australian recordkeeping Metadata standard, *Fifth Australian World Wide Web conference*. Southern Cross University, Australia.

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

• Miller, G. (1995). WordNet: A Lexical Database for English, *COMMUNICATIONS OF THE ACM*, 38(11), pp.39-41.

• Motik, B., Maedche, A., & Volz, R. (2005). Ontology Representation and Querying for Realizing Semantics-Driven Applications. *G. Stamou, S. Kollias (Eds.): Multimedia Content and the Semantic Web: Methods, Standards and Tools*, 45-73.

• NISCO press (2004). *Understanding Metadata*. National Information Standards Organization, USA.

• Olfat, H. (2013). *Automatic Spatial Metadata Updating and Enrichment*. (unpublished Doctor of philosophy thesis), The University of Melbourne – infrastructure Engineering school, Australia.

• Parankusham, K. & Madupu, R. (2006). *Role of Metadata in the data warehousing Environments*. (unpublished Master thesis), Lulea University of Technology, Sweden.

• Potok, T. E., Elmore, M. T., Reed, J. W., & Samatova, N. F. (2002, January). An ontology-based HTML to XML conversion using intelligent agents. In *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on* (pp. 1220-1229). IEEE..

• Ribeiro, M. (2014). *Extraction of non-taxonomic relations from texts to enrich a basic ontology*, Universidade Tecnica de Lisboa, portogal .

• Rodríguez, H., Farwell, D., Farreres, J., Bertran, M., Alkhalifa, M., Martí, M. A., ... & Vossen, P. (2008, January). Arabic wordnet: Current state and future extensions. In *Proceedings of The Fourth Global WordNet Conference, Szeged, Hungary*..

- Roussey, C. Pient, F. AH Kang, M. Corcho, O. (2011). An Introduction to Ontologies and Ontology Engineering, in book: *Ontologies in Urban Development Projects*, Springer, pp.9-38

- Shaheen, A. Al-Sayyed, R. Sleit, A. (2017). Improving visual analyses and communications of ontology by dynamic tree (case study: computer system), *International Journal of Advanced and Applied Science*, 4(5), pp. 62-66

- Sicilia, M. (2006). Metadata, semantics, and ontology: providing meaning to information resources. *Int. J. Metadata, Semantics and Ontologies*, 1(1): pp 83-86.

- Sini, M., Salokhe, G., Pardy, C., Albert, J., Keizer, J. and Katz, S. (2007). Ontology-based navigation of bibliographic metadata: example of the Food, Nutrition and Agriculture Journal. *In Proceedings of the International Conference on the Semantic Web and Digital Libraries*, DRTC, Bangalore.

- Sistrunk, W. (2002). Metadata*:* An overview with examples. *In: Missouri Digitization Conference*, Columbia, (Online) available : http://s.web.umkc.edu/sistrunkw/MLNC.htm.

- Sugumaran, Vijayan (2012). *Organizational efficiency through intelligent information technologies*, IGI Global.

- Teng, Y., Du, W., López-Ortiz, A., & Coleman, D. J. (2000). *Use of xml for web-based query processing of geospatial data*. University of New Brunswick (Canada).

- Winder, P., Eisenhauer, G. & Schwan, K. (2002). "*Open Metadata Formats: Efficient XML-Based Communication for High Performance Computing*". Cluster Computing, 5(3), pp. 315–324.

Appendix

Level two tables:

Semi-Manual Extracted Terms

اجتماعات	استجابة	استعمال	اسم
اعتماد	امتحان	إجابة	إجراء – إجراءات
إداري	إعادة	إقرار-إقرارها	إنجليزية
أداء	أردني	أستاذ	أعضاء
أقسام	أكاديمي	أمناء	أهمية
أوافق	باحث – باحثون	بحث -بحوث	بند
بيانات	تُعرض	تُلغى	تاريخ
تبني	تحقق	تخصص	تدريسية
ترفق	ترقية	تسجيل	تطبيق
تطويرها	تعديل-تعديلات	تعرض	تعليمات
تعيين	تقرير	تقوم	تقييم
تلغى	تنفيذ	توصية-توصيات	توقيع
جامعة	جامعي	جلسة	جهات
جودة – نوعية	خارجي	خطة	دراسات-دراسة
درجة	دعم	دقيق	دينار
رأي	رتبة	رسوم	رفع
رقم	رؤية	رئيس	زمن-زمنية
سابقة	ساعة-ساعات	سلامة	شروط
شهور	شؤون	صيغ-صياغة-صيغة	ضعيف
ضمان	طالب	طباعة	طريقة
عام	عامة	عربية	علامة
علمي-علمية	عمادة	عمل	عمل
عمليات	عميد-عمداء	عنوان	فصل
فقرة	فني	قائمة	قبول
قرار	قسم – دائرة	قياس	قيمة

كتاب	كفاءة	كلية-كليات	لأغراض
لجنة	لغة	ماجستير	مادة
مالية	متعلقة	مجلس - مجالس	مجموع
محاضر	محتوى-محتويات	محكم	مخطوط
مدة	مدير	مراجع	مركز-مراكز
مساعد	مساهمة	مستوى	مسؤولية
مشارك	مشروع	مطلوب	معدل-تعديلات
معلومات	مفردات	مقبول	مقدار
مقدر	مقرر	مكتب. مقر. مركز	ملاحظة
ملائمة	ملخص	مناقشة	منحة
مهمة – رسالة	مواد	موافقة	موضوع
مؤشر-مؤشرات	مؤلف	ناجح	نتيجة
نسخة	نشرة	نموذج-نماذج	نهائية
هاتف	هدف-أهداف	وثيقة	وضوح
يوم			

Concept extracting by KAON tool for thersold 5

academic	acceptable	accepted	adjustment
adjustments	admission	admissions	adoption
amended	amendments	answer	application
approval	Arabic	assistant	assurance
author	authorities	board	book
center	clarity	college	committee
contents	copy	council	date
dean	deanship	decision	department
dinars	director	directors	discussion
estimated	evaluation	exam	faculty
fee	final	form	forms
goal	graduate	grant	history
hours	implementation	importance	important
indication	indicator	information	item
manuscript	meetings	member	mission
model	modification	months	name
note	number	objectives	office
operations	opinion	paragraph	procedure
professor	project	quality	raise
recommendation	registration	related	report
required	research	researcher	response
responsibility	result	scientific	section
sections	sections	session	set
signature	specialization	student	studies
study	support	time	title
use	used	total	university
vision	written	value	version
year			

Comparison over both concepts extraction list threshold 5

	English concepts extraction by KAON tool	Translated concepts to Arabic by Google translate	Arabic concepts extraction semi-manually	Matching
1.	academic	أكاديمي	أكاديمي	Match
2.	acceptable	مقبول	مقبول	Match
3.	accepted	قبلت	مقبول	Match
4.	adjustment-adjustments	تعديل	تعديل-تعديلات	Match
5.	admission-admissions	قبول	القبول	Match
6.	adoption	تبني	تبني	Match
7.	amended-amendments	معدل-تعديلات	معدل-تعديلات	Match
8.	answer	إجابة	إجابة	Match
9.	application	الوضعية	التطبيق	Not Match
10.	approval	موافقة	الموافقة	Match
11.	Arabic	عربي	العربية	Match
12.	assistant	مساعد	مساعد	Match
13.	assurance	توكيد	ضمان	Similar
14.	author	مؤلف	المؤلف	Match
15.	authorities	السلطات	الجهات	Similar
16.	book	كتاب	الكتاب	Match
17.	center	مركز	المركز	Match
18.	clarity	وضوح	الوضوح	Match
19.	college	كلية	الكلية	Match
20.	committee	لجنة	لجنة	Match
21.	contents	محتويات	محتويات	Match
22.	copy	نسخ	نسخة	Match

Comparison over both concepts extraction list threshold 5

23.	council	مجلس	مجلس	Match
24.	date	تاريخ	تاريخ	Match
25.	dean	عميد	عميد	Match
26.	deanship	عمادة	عمادة	Match
27.	decision	قرار	قرار	Match
28.	department	قسم، أقسام ⁇	القسم-الدائرة	Match
29.	dinars	دينار	دينار	Match
30.	director	مدير	مدير	Match
31.	directors	الإدارة		Not Match
32.	discussion	نقاش	المناقشة	Match
33.	estimated	مقدر	المقدر	Match
34.	evaluation	تقييم	تقييم	Match
35.	exam	امتحان	الامتحان	Match
36.	faculty	كلية	كلية-كليات	Match
37.	fee	رسوم	رسوم	Match
38.	final	نهائي	النهائية	Match
39.	form-forms	استمارات	نموذج-النماذج	Similar
40.	goal	هدف	الهدف	Match
41.	graduate	تخرج		Not Match
42.	grant	منحة	المنحة	Match
43.	hours	ساعات	الساعات	Match
44.	implementation	التنفيذ	تنفيذ	Match
45.	importance	أهمية	أهمية	Match
46.	important	مهم		Not Match
47.	indication	إشارة	علامة	Similar
48.	indicator	مؤشر	المؤشر	Match

49.	information	معلومات	المعلومات	Match
50.	item	بند	البند	Match
51.	manuscript	مخطوطة	المخطوط	Match
52.	meetings	اجتماعات	اجتماعات	Match
53.	member	عضو	أعضاء	Similar
54.	mission	مهمة	مهمة - رسالة	Match
55.	model	نموذج	نموذج	Match
56.	modification	تعديل	تعديل	Match
57.	months	الشهور	شهور	Match
58.	name	اسم	الاسم	Match
59.	note	ملحوظة	ملاحظة	Match
60.	number	رقم	الرقم	Match
61.	objectives	الأهداف	الأهداف	Match
62.	office	مكتب. مقر. مركز	مكتب. مقر. مركز	Match
63.	operations	عمليات	العمليات	Match
64.	opinion	رأي	رأي	Match
65.	paragraph	فقرة	فقرة	Match
66.	procedure	إجراء	إجراء	Match
67.	professor	دكتور جامعي	أستاذ	Match
68.	project	مشروع	مشروع	Match
69.	quality	جودة	الجودة - النوعية	Match
70.	raise	ربى	رفع	Not Match
71.	recommendation	توصية	توصية	Match
72.	registration	التسجيل	التسجيل	Match
73.	related	ذات صلة	المتعلقة	Similar
74.	report	أبلغ عن	تقرير	Similar

75.	required	مطلوب	المطلوب	Match
76.	research	ابحاث	البحوث	Match
77.	researcher	الباحث	الباحث	Match
78.	response	استجابة	الاستجابة	Match
79.	responsibility	المسئولية	مسؤولية	Match
80.	result	نتيجة	نتيجة	Match
81.	scientific	علمي	العلمي	Match
82.	section	الجزء	قسم – دائرة	Not Match
83.	sections	الأقسام	الأقسام	Match
84.	session	جلسة	الجلسة	Match
85.	set	جلس		Not Match
86.	signature	التوقيع	التوقيع	Match
87.	specialization	تخصص	التخصص	Match
88.	student	طالب علم	الطالب	Match
89.	studies-study	دراسات-دراسة	دراسات-دراسة	Match
90.	support	الدعم	الدعم	Match
91.	time	زمن	الزمن	Match
92.	title	عنوان	عنوان	Match
93.	total	مجموع	المجموع	Match
94.	university	جامعة	الجامعة	Match
95.	use	استعمال	استعمال	Match
96.	used	مستخدم		Not Match
97.	value	القيمة	القيمة	Match
98.	version	الإصدار		Not Match
99.	vision	رؤية	الرؤية	Match
100.	written	مكتوب		Not Match
101.	year	عام	العام	Match

Concept extracting by KAON tool for thersold 3

academic	acceptable	accepted	action
address	adjustment	adjustments	administrative
admission	admissions	adoption	amended
amendments	amount	answer	Arabic
application	appropriate	approval	assurance
arbitrator	author	assistant	board
attached	canceled	authorities	chapter
book	college	center	conditions
clarity	contribution	committee	council
contents	date	copy	dean
data	decision	day	department
deanship	dinars	degree	directors
development	display	director	document
discussion	English	distance	evaluation
efficiency	exam	estimated	external
event	fee	extent	form
faculty	formulation	final	good
forms	grant	goal	history
graduate	implementation	head	important
hours	indicator	importance	instructions
indication	Jordanian	information	level
item	list	language	master
light	means	manuscript	measures
material	member	measure	method
meetings	model	message	months

mission	note	modification	objectives
name	operations	number	other
office	participant	opinion	period
paragraph	plan	performance	procedure
phone	project	print	publication
professor	quality	public	recommendation
purposes	registration	raise	report
references	research	related	response
required	result	researcher	scientific
responsibility	second	safety	sections
search	session	section	signature
semester	specialization	set	studies
signing	successful	student	support
study	technician	summary	terms
target	title	telephone	transfer
time	university	total	used
trustees	version	use	vocabulary
value	work	vision	written
weak	working	year	

Comparison over both concepts extraction list threshold 3

	English concepts extraction by KAON tool	Translated concepts to Arabic by Google translate	Arabic concepts extraction semi-manually	Matching
1.	academic	أكاديمي	أكاديمي	Match
2.	acceptable	مقبول	مقبول	Match
3.	accepted	قبلت	مقبول	Match
4.	action	عمل	عمل	Match
5.	address	عنوان	عنوان	Match
6.	adjustment-adjustments	تعديل	تعديل-تعديلات	Match
7.	administrative	إداري	الإداري	Match
8.	admission-admissions	قبول	القبول	Match
9.	adoption	تبني	تبني	Match
10.	amended-amendments	معدل-تعديلات	معدل-تعديلات	Match
11.	amount	كمية	مقدار	Similar
12.	answer	إجابة	إجابة	Match
13.	application	الوضعية	التطبيق	Not Match
14.	appropriate	مناسب	ملائمة	Match
15.	approval	موافقة	الموافقة	Match
16.	Arabic	عربي	العربية	Match
17.	arbitrator	المحكم	المحكم	Match
18.	assistant	مساعد	مساعد	Match
19.	assurance	توكيد	ضمان	Similar
20.	attached	تعلق	ترفق	Similar
21.	author	مؤلف	المؤلف	Match
22.	authorities	السلطات	الجهات	Similar

23.	board	مجلس	مجلس	Match
24.	book	كتاب	الكتاب	Match
25.	canceled	ألغيت	تلغى	Match
26.	center	مركز	المركز	Match
27.	chapter	الفصل	الفصل	Match
28.	clarity	وضوح	الوضوح	Match
29.	college	كلية	الكلية	Match
30.	committee	لجنة	لجنة	Match
31.	conditions	الظروف	شروط	Similar
32.	contents	محتويات	محتويات	Match
33.	contribution	إسهام	مساهمة	Match
34.	copy	نسخ	نسخة	Match
35.	council	مجلس	مجلس	Match
36.	data	البيانات	البيانات	Match
37.	date	تاريخ	تاريخ	Match
38.	day	يوم	يوم	Match
39.	dean	عميد	عميد	Match
40.	deanship	عمادة	عمادة	Similar
41.	decision	قرار	قرار	Match
42.	degree	الدرجة العلمية	درجة	Match
43.	department	قسم، أقسام ⬚	القسم-الدائرة	Match
44.	development	تطوير	تطويرها	Match
45.	dinars	دينار	دينار	Match
46.	director	مدير	مدير	Match
47.	directors	الإدارة		Not Match
48.	discussion	نقاش	المناقشة	Match

49.	display	عرض	تعرض	Match
50.	distance	مسافه: بعد		Not Match
51.	document	وثيقة	وثيقة	Match
52.	efficiency	نجاعة	الكفاءة	Not Match
53.	English	الإنجليزية	الإنجليزية	Match
54.	estimated	مقدر	المقدر	Match
55.	evaluation	تقييم	تقييم	Match
56.	event	هدف	الهدف	Match
57.	exam	امتحان	الامتحان	Match
58.	extent	مدى		Not Match
59.	external	خارجي	الخارجي	Match
60.	faculty	كلية	الكلية	Match
61.	fee	رسوم	رسوم	Match
62.	final	نهائي	النهائية	Match
63.	form-forms	استمارات	نموذج-النماذج	Similar
64.	formulation	صيغة	الصيغة	Match
65.	goal	هدف	الهدف	Match
66.	good	جيد		Not Match
67.	graduate	تخرج		Not Match
68.	grant	منحة	المنحة	Match
69.	head	رئيس	رئيس	Match
70.	history	التاريخ	تاريخ	Match
71.	hours	ساعات	الساعات	Match
72.	implementation	التنفيذ	تنفيذ	Match
73.	importance	أهمية	أهمية	Match
74.	important	مهم	أهمية	Similar

75.	indication	إشارة	علامة	Similar
76.	indicator	مؤشر	المؤشر	Match
77.	information	معلومات	المعلومات	Match
78.	instructions	تعليمات	التعليمات	Match
79.	item	بند	البند	Match
80.	Jordanian	أردني	الأردني	Match
81.	language	لغة	اللغة	Match
82.	level	مستوى	مستوى	Match
83.	light	ضوء		Not Match
84.	list	قائمة	قائمة	Match
85.	manuscript	مخطوطة	المخطوط	Match
86.	master	رئيسي - سيد	ماجستير	Not Match
87.	material	مواد	المواد	Match
88.	means	يعني		Not Match
89.	measure	قياس	قياس	Match
90.	measures	الإجراءات	الإجراءات	Match
91.	meetings	اجتماعات	اجتماعات	Match
92.	member	عضو	أعضاء	Similar
93.	message	رسالة	الرسالة	Match
94.	method	طريقة	طريقة	Match
95.	mission	مهمة	مهمة - رسالة	Match
96.	model	نموذج	نموذج	Match
97.	modification	تعديل	تعديل	Match
98.	months	شهور	شهور	Match
99.	name	اسم	الاسم	Match
100.	note	ملحوظة	ملاحظة	Match

101.	number	رقم	الرقم	Match
102.	objectives	الأهداف	الأهداف	Match
103.	office	مكتب. مقر. مركز	مكتب. مقر. مركز	Match
104.	operations	عمليات	العمليات	Match
105.	opinion	رأي	رأي	Match
106.	other	آخر		Not Match
107.	paragraph	فقرة	فقرة	Match
108.	participant	مشارك	مشارك	Match
109.	performance	أداء	الأداء	Match
110.	period	فترة	المدة	Match
111.	phone	هاتف	الهاتف	Match
112.	plan	خطة	الخطة	Match
113.	print	طباعة	طباعة	Match
114.	procedure	إجراء	إجراء	Match
115.	professor	دكتور جامعي	أستاذ	Match
116.	project	مشروع	مشروع	Match
117.	public	عامة	العامة	Match
118.	publication	منشور	نشرة	Similar
119.	purposes	المقاصد	الهدف	Similar
120.	quality	جودة	الجودة ـ النوعية	Match
121.	raise	ربى	رفع	Not Match
122.	recommendation	توصية	توصية	Match
123.	references	المراجع	المراجع	Match
124.	registration	التسجيل	التسجيل	Match
125.	related	ذات صلة	المتعلقة	Similar
126.	report	أبلغ عن	تقرير	Similar

127.	required	مطلوب	المطلوب	Match
128.	research	ابحاث	البحوث	Match
129.	researcher	الباحث	الباحث	Match
130.	response	استجابة	الاستجابة	Match
131.	responsibility	المسئولية	مسؤولية	Match
132.	result	نتيجة	نتيجة	Match
133.	safety	سلامة	سلامة	Match
134.	scientific	علمي	العلمي	Match
135.	search	بحث	البحث	Match
136.	second	ثانيا		Not Match
137.	section	الجزء	قسم – دائرة	Not Match
138.	sections	الأقسام	الأقسام	Match
139.	semester	نصف السنة		Not Match
140.	session	جلسة	الجلسة	Match
141.	set	جلس		Not Match
142.	signature	التوقيع	التوقيع	Match
143.	signing	التوقيع	التوقيع	Match
144.	specialization	تخصص	التخصص	Match
145.	student	طالب علم	الطالب	Match
146.	studies-study	دراسات-دراسة	دراسات-دراسة	Match
147.	successful	ناجح	ناجح	Match
148.	summary	ملخص	ملخص	Match
149.	support	الدعم	الدعم	Match
150.	target	استهداف	الهدف	Match
151.	technician	فني	الفني	Match
152.	telephone	هاتف	الهاتف	Match

153.	terms	شروط	شروط	Match
154.	time	زمن	الزمن	Match
155.	title	عنوان	عنوان	Match
156.	total	مجموع	المجموع	Match
157.	transfer	تحويل		Not Match
158.	trustees	الأمناء	الأمناء	Match
159.	university	جامعة	الجامعة	Match
160.	use	استعمال	استعمال	Match
161.	used	مستخدم		Not Match
162.	value	القيمة	القيمة	Match
163.	version	الإصدار		Not Match
164.	vision	رؤية	الرؤية	Match
165.	vocabulary	مفردات اللغه	مفردات	Match
166.	weak	ضعيف	ضعيف	Match
167.	work	عمل	العمل	Match
168.	working	عامل		Not Match
169.	written	مكتوب		Not Match
170.	year	عام	العام	Match

Level Three tables:

Semi-Manual Extracted Terms

اجتماعات	اجتهاد	اختبارات	اختياري
استبانة	استجابة	استطلاع	استعمال
اسم	اعتماد	اقتراحات	اقرار
الاكتروني	التحاق	التزام	التسديد
الطالب	اللازمة	اللغة	المالي - المالية
المعنية	امتحان	انترنت	انذار
انسحاب	اهتمام	إجابة	إجراء - إجراءات
إخطار	إدارة	إداري	إشارة
إشعار	إعادة	إعداد	إقرار
إنتاجية	إنجليزية	إنذار	أتعهد
أجهزة	أخرى	أخلاق	أداء
أردني	أسباب	أستاذ	أعضاء
أقسام	أكاديمي	أمناء	أنظمة
أهمية	أوافق	باحث - باحثون	بحث -بحوث
برنامج	بريد	بشرية	بشكل
بكالوريوس	بند	بيان	بيانات
تُعرض	تُلغى	تاريخ	تأديبي
تبني	تتوافر	تجهيزات	تحقق
تحقيق	تحميل	تخرج	تخصص
تخفيض	تدريب- تدريبية	تدريسية	تدقيق
تراكمي	ترفق	ترقية	تسجيل
تصميم	تطبيق - تطبيقات	تطوير	تعاميم
تعديل-تعديلات	تعريف	تعليمات	تعليمية
تعيين	تقدير	تقديم	تقرير-تقارير
تقيد	تقييم	تكنولوجيا	تلغى

تنال	تنسيقات	تنظيم	تنظيمي
تنفيذ	تنقل	توافق	توصية-توصيات
توفر	توقيع	ثباتة	جامعة
جامعي	جلسة	جنسية	جهة-جهات
جودة - نوعية	جيد	حديثة	حساب
حصول	خارجي	خدمة-خدمات	خطة
داخلي	دخول	دراسة-دراسات	دراسي
درجة	دعم	دقيق	دورة
دوريات	دينار	رأي	رتبة
رسوم	رفع	رقابة	رقم
رمز	رؤية	رئيس	زمن-زمنية
سابقة	ساعة-ساعات	سرية	سريعة
سلامة	سهولة	شبكة	شرق
شروط	شهرياً-شهور	شؤون	صفحات
صيغ-صياغة-صيغة	صيفي	ضعيف	ضمان
طباعة	طريقة	طلبة	عام
عامة	عاملون-عاملين	عربية	عرض
علاقة	علمي-علمية	عمادة	عمل
عملية-عمليات	عميد-عمداء	عنوان	فصل-فصول
فعالية	فقرات	فني	قاعة
قائمة	قبول	قرار-قرارات	قسم - دائرة
قليلة	قياس	قيمة	كافي-كافية
كاملة	كتاب	كفاءة	كلية-كليات
لأغراض	لجنة	ماجستير	مادة
مبررات	مبينة	متعلقة	متميز
متوسطة	مثالاً	مجال	مجلس-مجالس
مجموع	محاضر	محتوى-محتويات	محكم

مختارة	مخطوط	مدة	مدرب
مدير	مراجع	مرفق	مركز-مراكز
مساعد	مساعدة	مساهمة	مستخدم
مستوى	مسجل	مسؤولية	مشارك-مشاركين
مشرف	مشروع	مصادر	مطابقة
مطلوب	معتمدة	معدل	معرفة
معلومات	مفردات	مقبول	مقدار
مقدر	مقدمة	مقرر	مكان
مكتب. مقر. مركز	مكتبة	مكتوب	ملائمة
ملحوظة	ملخص	ملفات	ممتاز
مناسب	مناسبة	مناقشة	منحة
منشورة	منظمة	مهمة ـ رسالة	مواد
موارد-مصادر	مواظبة	مواعيد	موافقة
موضوع	موظف	موعد	موقع
مؤشر-مؤشرات	مؤلف	ناجح	نتيجة
نسبة	نسخة	نشاط	نظام
نموذج-نماذج	نهاية	نهائي	هاتف
هدف-أهداف	هوية	هيكل	هيئة
وثيقة	وحدة	ورشة	وسائل
وضوح	يُلغى	يتحمل	يستثنى
يسهل	يكتب	يوم	

Concept extracting by KAON tool threshold 5

absence	academic	accepted	access
achieved	action	activity	adjustment-adjustments
administrative	admission-admissions	adoption	amended-amendments
amount	answer	application	applications
approval	Arabic	assessed-assessment-assistant	assurance
author	authorities	board	book
center	chapter	clarity	coach
college	commitment	committee	compatibility
condition	content-contents	copy	council
course	data	date	day
dean	deanship	decision	decisions
degree	department-departments	development	dinars
director	discussion	document	east
efficiency	employee	evaluation	event
exam	facility	faculty	fee
field	file	form-forms	goal
grant	head	history	hours
human	identity	implementation	indicator-indicators
information	instructions	item	language
level	library	management	manager
master	material	means	means

meetings	member	message	mission
model	modern	modification	month-months
move	name	needs	note
notice	number	objection	objectives
office	operations	opinion	organization
organizer	page	paragraph	participant
performance	period	periodicals	place
plan	presentation	print	procedure-procedures
process	professor	project	publication
purposes	quality	questionnaire	Raise
reasons	recommendation-recommendations	registration	regulations
report	requirements	research	researcher
resolution	resources	response	responsibility
result	safety	section	semester
service	session	signature	site
sources	specialization	staff	student
studies	study	supervisor	support
system	technology	telephone	terms
time	timing	title	training
trustees	university	use	value
version	vision	vocabulary	warning
work	workshop	year	

Comparison over both concepts extraction list threshold 5

	English concepts extraction by KAON tool	Translated concepts to Arabic by Google translate	Arabic concepts extraction semi-manually	Matching
1.	absence	غياب		Not Match
2.	academic	أكاديمي	أكاديمي	Match
3.	accepted	قبلت	مقبول	Match
4.	access	التمكن من	الدخول	Match
5.	achieved	حقق	تحقق	Match
6.	action	عمل	عمل	Match
7.	activity	نشاط	نشاط	Match
8.	adjustment-adjustments	تعديل	تعديل-التعديلات	Match
9.	administrative	إداري	الإداري	Match
10.	admission-admissions	قبول	القبول	Match
11.	adoption	تبني	تبني	Match
12.	amended	معدل	معدل	Match
13.	amendments	معدل	تعديلات	Match
14.	amount	كمية	مقدار	Similar
15.	answer	إجابة	إجابة	Match
16.	application	الوضعية	التطبيق	Not Match
17.	applications	تطبيقات	التطبيقات	Match
18.	approval	موافقة	موافقة	Match
19.	Arabic	عربي	العربية	Match
20.	assistant	مساعدة	مساعدة	Match
21.	assessed-assessment	تقييم-تقدير	تقييم-تقدير	Match
22.	assurance	توكيد	ضمان	Similar

23.	author	مؤلف	المؤلف	Match
24.	authorities	السلطات	الجهات	Similar
25.	board	مجلس	مجلس	Match
26.	book	كتاب	الكتاب	Match
27.	center	مركز	مركز	Match
28.	chapter	الفصل	الفصل	Match
29.	clarity	وضوح	وضوح	Match
30.	coach	مدرب حافلة ركاب	المدرب	Match
31.	college	كلية	الكلية	Match
32.	commitment	التزام	التزام	Match
33.	committee	لجنة	اللجنة	Match
34.	compatibility	التوافق	التوافق	Match
35.	condition	الظروف	شروط	Similar
36.	content-contents	يحتوى	محتوى -محتويات	Match
37.	copy	نسخ	نسخة	Match
38.	council	مجلس	مجلس	Match
39.	course	دورة	دورة	Match
40.	data	البيانات	البيانات	Match
41.	date	تاريخ	تاريخ	Match
42.	day	يوم	يوم	Match
43.	dean	عميد	عميد	Match
44.	deanship	عمادة	عمادة	Match
45.	decision	قرار	قرار	Match
46.	decisions	قرارات	القرارات	Match
47.	degree	الدرجة العلمية	درجة	Match
48.	department-	قسم-دائرة	القسم-الدائرة-إدارة	Match

	departments			
49.	development	تطوير	تطوير	Match
50.	dinars	دينار	دينار	Match
51.	director	مدير	مدير	Match
52.	discussion	نقاش	المناقشة	Match
53.	document	وثيقة	وثيقة	Match
54.	east	الشرق	الشرق	Match
55.	efficiency	نجاعة	الكفاءة	Not Match
56.	employee	موظف	موظف	Match
57.	evaluation	تقييم	تقييم	Match
58.	event	هدف	الهدف	Match
59.	exam	امتحان	الامتحان	Match
60.	facility	منشأة	مرفق	Similar
61.	faculty	كلية	الكلية	Match
62.	fee	رسوم	رسوم	Match
63.	field	حقل	مجال	Similar
64.	file	ملف	الملفات	Match
65.	form	نموذج	نموذج	Match
66.	forms	استمارات	نموذج-النماذج	Similar
67.	goal	هدف	الهدف	Match
68.	grant	منحة	المنحة	Match
69.	head	رئيس	رئيس	Match
70.	history	التاريخ	تاريخ	Match
71.	hours	ساعات	ساعات	Match
72.	human	بشري	بشرية	Match
73.	identity	هوية	هوية	Match

74.	implementation	التنفيذ	تنفيذ	Match
75.	indicator	مؤشر	المؤشر-مؤشرات	Match
76.	indicators	مؤشرات	المؤشر-مؤشرات	Match
77.	information	معلومات	المعلومات	Match
78.	instructions	تعليمات	التعليمات	Match
79.	item	بند	بند	Match
80.	language	لغة	اللغة	Match
81.	level	مستوى	مستوى	Match
82.	library	مكتبة	مكتبة	Match
83.	management	إدارة	إدارة	Match
84.	manager	مدير	مدير	Match
85.	master	رئيسي - سيد	ماجستير	Not Match
86.	material	مواد	المواد	Match
87.	means	يعني		Not Match
88.	meetings	اجتماعات	اجتماعات	Match
89.	member	عضو	أعضاء	Similar
90.	message	رسالة	الرسالة	Match
91.	mission	مهمة	مهمة - رسالة	Match
92.	model	نموذج	نموذج	Match
93.	modern	حديث	الحديثة	Similar
94.	modification	تعديل	تعديل	Match
95.	month-months	شهر	شهرياً-شهور	Match
96.	move	نقل	تنقل	Similar
97.	name	اسم	الاسم	Match
98.	needs	الاحتياجات	الاحتياجات	Match
99.	note	ملحوظة	ملاحظة	Match

100.	notice	تنويه	إشعار	Match
101.	number	رقم	الرقم	Match
102.	objection	اعتراض		Not Match
103.	objectives	الأهداف	الأهداف	Match
104.	office	مكتب. مقر. مركز	مكتب. مقر. مركز	Match
105.	operations	عمليات	عمليات	Match
106.	opinion	رأي	رأي	Match
107.	organization	منظمة	منظمة	Match
108.	organizer	منظم		Not Match
109.	page	صفحة	صفحات	Match
110.	paragraph	فقرة	الفقرات	Similar
111.	participant	مشارك	مشارك	Match
112.	performance	أداء	الأداء	Match
113.	period	فترة	المدة	Similar
114.	periodicals	الدوريات	الدوريات	Match
115.	place	مكان	مكان	Match
116.	plan	خطة	الخطة	Match
117.	presentation	عرض	عرض	Match
118.	print	طباعة	طباعة	Match
119.	procedure-procedures	إجراء	إجراء-إجراءات	Match
120.	process	معالج	عملية	Similar
121.	professor	دكتور جامعي	أستاذ	Match
122.	project	مشروع	مشروع	Match
123.	publication	منشور	المنشورة	Similar
124.	purposes	المقاصد	الهدف	Similar
125.	quality	جودة	الجودة - النوعية	Match

126.	questionnaire	استطلاع	استطلاع	Match
127.	Raise	ربى	رفع	Not Match
128.	reasons	أسباب	أسباب	Match
129.	recommendation-recommendations	توصية	توصية-توصيات	Match
130.	registration	التسجيل	التسجيل	Match
131.	regulations	قوانين	أنظمة	Similar
132.	report	أبلغ عن	تقرير	Similar
133.	requirements	المتطلبات	متطلبات	Match
134.	research	ابحاث	البحوث	Match
135.	researcher	الباحث	الباحث	Match
136.	resolution	القرار	قرار	Match
137.	resources	موارد	موارد-مصادر	Match
138.	response	استجابة	الاستجابة	Match
139.	responsibility	المسئولية	مسؤولية	Match
140.	result	نتيجة	نتيجة	Match
141.	safety	سلامة	سلامة	Match
142.	section	الجزء	قسم - دائرة	Not Match
143.	semester	نصف السنة		Not Match
144.	service	الخدمات	خدمات	Match
145.	session	جلسة	الجلسة	Match
146.	signature	التوقيع	توقيع	Match
147.	site	موقع	موقع	Match
148.	sources	مصادر	مصادر	Match
149.	specialization	تخصص	التخصص	Match
150.	staff	العاملين	العاملين	Match

151.	student	طالب علم	الطالب	Match
152.	studies	دراسات	الدراسات	Match
153.	study	دراسة	دراسة	Match
154.	supervisor	مشرف	مشرف	Match
155.	support	الدعم	الدعم	Match
156.	system	النظام	نظام	Match
157.	technology	تقنية	تكنولوجيا	Similar
158.	telephone	هاتف	هاتف	Match
159.	terms	شروط	شروط	Match
160.	time	زمن	الزمن	Match
161.	timing	توقيت	توقيت	Match
162.	title	عنوان	عنوان	Match
163.	training	تدريب	تدريب	Match
164.	trustees	الأمناء	الأمناء	Match
165.	university	جامعة	الجامعة	Match
166.	use	استعمال	استعمال	Match
167.	value	القيمة	القيمة	Match
168.	version	الإصدار		Not Match
169.	vision	رؤية	الرؤية	Match
170.	vocabulary	مفردات اللغة	مفردات	Match
171.	warning	تحذير	انذار	Match
172.	work	عمل	العمل	Match
173.	workshop	ورشة عمل	ورشة	Match
174.	year	عام	العام	Match

Concept extracting by KAON tool threshold 3

absence	academic	acceptable	accepted
access	account	achieved	action
activity	address	adequate	adjustment-adjustments
administrative	admission-admissions	adoption	affairs
allocated	amended-amendments	amount	answer
application	applications	appropriate	approval
approving	Arabic	arbitrator	assessed-assessment-assistant
assurance	attached	attention	audit
author	authorities	availability	average
bachelor	board	book	canceled
center	chapter	chief	circulars
clarity	classes	classroom	coach
collection	college	commitment	committee
compatibility	comprehensive	concerned	condition
confidentiality	content-contents	contribution	control
copy	council	course	cumulative
data	date	day	dean
deanship	decision	definition	degree
department-departments	design	development	devices
diligent	dinars	director	directors
disciplinary	discussion	display	distance

document	duration	east	education
effectiveness	efficiency	electronic	email
employee	end	English	estimated
ethics	evaluation	event	exam
example	excellent	extent	external
facility	faculty	fee	field
file	final	financial	find
fixed	following	force	form-forms
formulation	gain	goal	good
graduate	grant	hall	head
history	hours	human	id
ident	identical	identity	image
implementation	importance	important	indication
indicator-indicators	information	instructions	internal
internet	item	items	joining
Jordanian	justification	language	level
library	light	limits	list
loading	location	magazine	mail
management	manager	manner	manuscript
master	material	meaning	means
measure	measures	meetings	member
message	method	mission	model
modern	modification	month-months	move
name	nationality	necessary	need
network	note	notice	number
objection	objectives	obtained	office

operations	opinion	option	organization
organizer	other	outstanding	page
paragraph	participant	participants	pay
performance	period	periodicals	perseverance
phone	place	plan	pledge
preparation	prescribed	presentation	president
previous	print	procedure-procedures	process
productivity	professor	program	project
proportion	proposals	provision	provisions
public	publication	purpose	purposes
quadrilateral	quality	questionnaire	Raise
rate	reasons	recommendation-recommendations	reduced
reference	registrar	registration	regulations
related	relevance	relevant	report
reports	request	required	research
researcher	resolution	resources	response
responsibility	result	safety	scientific
search	second	section	sections
selected	semester	service	session
set	signature	signing	site
sources	specialization	staff	student
studies	study	subject	successful
successive	sufficient	suggestions	suitable
summary	summer	supervisor	support
system	target	targets	teaching

teams	technician	technology	telephone
terms	tests	text	time
title	topic	total	training
transfer	treated	trustees	type
unit	university	use	used
value	version	vision	vocabulary
warning	weak	weakness	whole
withdrawal	work	working	workplace
workshop	written	year	

Comparison over both concepts extraction list threshold 3

	English concepts extraction by KAON tool	Translated concepts to Arabic by Google translate	Arabic concepts extraction semi-manually	Matching
1.	absence	غياب		Not Match
2.	academic	أكاديمي	أكاديمي	Match
3.	acceptable	مقبول	مقبول	Match
4.	accepted	قبلت	مقبول	Similar
5.	access	التمكن من	دخول	Match
6.	account	الحساب	حساب	Match
7.	achieved	حقق	تحقق	Match
8.	action	عمل	عمل	Match
9.	activity	نشاط	نشاط	Match
10.	address	عنوان	عنوان	Match
11.	adequate	كاف	كافية	Match
12.	adjustment-adjustments	تعديل	تعديل-التعديلات	Match
13.	administrative	إداري	الإداري	Match
14.	admission-admissions	قبول	القبول	Match
15.	adoption	تبني	تبني	Match
16.	affairs	أمور	شؤون	Similar
17.	allocated	تخصيص	تخصص	Similar
18.	amended-amendments	معدل	معدل-تعديلات	Match
19.	amount	كمية	مقدر	Similar
20.	answer	إجابة	إجابة	Match
21.	application	الوضعية	تطبيق	Not Match

22.	applications	تطبيقات	تطبيقات	Match
23.	appropriate	مناسب	مناسب	Match
24.	approval	موافقة	موافقة	Match
25.	approving	تصديق	اقرار	Similar
26.	Arabic	عربى	العربية	Match
27.	arbitrator	المحكم	المحكم	Match
28.	assessed-assessment-assistant	تقييم-تقدير-مساعد	تقييم-تقدير-مساعد	Match
29.	assurance	توكيد	ضمان	Similar
30.	attached	تعلق	مرفق	Similar
31.	attention	انتباه	اهتمام	Similar
32.	audit	تدقيق	تدقيق	Match
33.	author	مؤلف	المؤلف	Match
34.	authorities	السلطات	الجهات	Similar
35.	availability	توفر	توفر	Match
36.	average	معدل	معدل	Match
37.	bachelor	أعزب	البكالوريوس	Not Match
38.	board	مجلس	مجلس	Match
39.	book	كتاب	الكتاب	Match
40.	canceled	ألغيت	تلغى	Match
41.	center	مركز	المركز	Match
42.	chapter	الفصل	الفصل	Match
43.	chief	رئيس	رئيس	Match
44.	circulars	التعاميم	التعاميم	Match
45.	clarity	وضوح	الوضوح	Match
46.	classes	فصول	فصول	Match

47.	classroom	قاعة الدراسة	فصل	Similar
48.	coach	مدرب حافلة ركاب	مدرب	Match
49.	collection	مجموعة		Not Match
50.	college	كلية	الكلية	Match
51.	commitment	التزام	التزام	Match
52.	committee	لجنة	لجنة	Match
53.	compatibility	التوافق	التوافق	Match
54.	comprehensive	شامل		Not Match
55.	concerned	المعنية	المعنية	Match
56.	condition	الظروف	شروط	Similar
57.	confidentiality	سرية	سرية	Match
58.	content-contents	يحتوى	محتوى -محتويات	Match
59.	contribution	إسهام	مساهمة	Similar
60.	control	مراقبة	رقابة	Similar
61.	copy	نسخ	نسخة	Match
62.	council	مجلس	مجلس	Match
63.	course	دورة	دورة	Match
64.	cumulative	تراكمي	تراكمي	Match
65.	data	البيانات	بيانات	Match
66.	date	تاريخ	تاريخ	Match
67.	day	يوم	يوم	Match
68.	dean	عميد	عميد	Match
69.	deanship	عمادة	عمادة	Match
70.	decision	قرار	قرار	Match
71.	definition	تعريف	تعريف	Match

72.	degree	الدرجة العلمية	درجة	Match
73.	department-departments	قسم-دائرة	القسم-الدائرة-إدارة	Match
74.	design	التصميم	تصميم	Match
75.	development	تطوير	التطوير	Match
76.	devices	الأجهزة	أجهزة	Match
77.	diligent	مجتهد	اجتهاد	Similar
78.	dinars	دينار	دينار	Match
79.	director	مدير	مدير	Match
80.	directors	الإدارة	إدارة	Match
81.	disciplinary	الانضباط		Not Match
82.	discussion	نقاش	المناقشة	Match
83.	display	عرض	العرض	Match
84.	distance	مسافه: بعد		Not Match
85.	document	وثيقة	وثيقة	Match
86.	duration	المدة الزمنية	مدة	Match
87.	east	الشرق	شرق	Match
88.	education	التعليم	تعليمية	Similar
89.	effectiveness	فعالية	فعالية	Match
90.	efficiency	نجاعة	الكفاءة	Not Match
91.	electronic	إلكتروني	الالكتروني	Match
92.	email	البريد الإلكتروني	بريد	Match
93.	employee	موظف	موظف	Match
94.	end	النهاية	نهاية	Match
95.	English	الإنجليزية	الإنجليزية	Match
96.	estimated	مقدر	المقدر	Match

97.	ethics	أخلاق	أخلاق	Match
98.	evaluation	تقييم	تقييم	Match
99.	event	هدف	الهدف	Match
100.	exam	امتحان	الامتحان	Match
101.	example	مثال	مثالاً	Similar
102.	excellent	ممتاز	ممتاز	Match
103.	extent	مدى		Not Match
104.	external	خارجي	الخارجي	Match
105.	facility	منشأة	مرفق	Similar
106.	faculty	كلية	الكلية	Match
107.	fee	رسوم	رسوم	Match
108.	field	حقل	مجال	Similar
109.	file	ملف	الملفات	Match
110.	final	نهائي	نهائي	Match
111.	financial	المالية	المالية	Match
112.	find	تجد		Not Match
113.	fixed	ثابت	ثباتة	Similar
114.	following	التالية		Not Match
115.	force	فرض		Not Match
116.	form-forms	استمارات	نموذج-النماذج	Similar
117.	formulation	صيغة	الصيغة	Match
118.	gain	ربح		Not Match
119.	goal	هدف	الهدف	Match
120.	good	جيد	جيد	Match
121.	graduate	تخرج	تخرج	Match

122.	grant	منحة	المنحة	Match
123.	hall	صالة	قاعة	Similar
124.	head	رئيس	رئيس	Match
125.	history	التاريخ	تاريخ	Not Match
126.	hours	ساعات	الساعات	Match
127.	human	بشري	بشرية	Match
128.	id	هوية شخصية	هوية	Match
129.	ident	الرمز	الرمز	Match
130.	identical	مطابق	مطابقة	Match
131.	identity	هوية	هوية	Match
132.	image	صورة		Not Match
133.	implementation	التنفيذ	تنفيذ	Match
134.	importance	أهمية	أهمية	Match
135.	important	مهم		Not Match
136.	indication	إشارة	إشارة	Match
137.	indicator-indicators	مؤشر	المؤشر-مؤشرات	Match
138.	information	معلومات	المعلومات	Match
139.	instructions	تعليمات	التعليمات	Match
140.	internal	داخلي	داخلي	Match
141.	internet	الإنترنت	انترنت	Match
142.	item	بند	البند	Match
143.	items	العناصر		Not Match
144.	joining	انضمام	التحاق	Match
145.	Jordanian	أردني	الأردني	Match
146.	justification	مبرر	المبررات	Similar

147.	language	لغة	اللغة	Match
148.	level	مستوى	مستوى	Match
149.	library	مكتبة	مكتبة	Match
150.	light	ضوء		Not Match
151.	limits	حدود		Not Match
152.	list	قائمة	قائمة	Match
153.	loading	جار التحميل	تحميل	Similar
154.	location	موقعك	موقع	Match
155.	magazine	مجلة		Not Match
156.	mail	بريد	بريد	Match
157.	management	إدارة	إدارة	Match
158.	manager	مدير	مدير	Match
159.	manner	أسلوب		Not Match
160.	manuscript	مخطوطة	المخطوط	Match
161.	master	رئيسي - سيد	ماجستير	Not Match
162.	material	مواد	المواد	Match
163.	meaning	المعنى		Not Match
164.	means	يعني		Not Match
165.	measure	قياس	قياس	Match
166.	measures	الإجراءات	الإجراءات	Match
167.	meetings	اجتماعات	اجتماعات	Match
168.	member	عضو	أعضاء	Similar
169.	message	رسالة	الرسالة	Match
170.	method	طريقة	طريقة	Match
171.	mission	مهمة	مهمة ـ رسالة	Match

172.	model	نموذج	نموذج	Match
173.	modern	حديث	الحديثة	Similar
174.	modification	تعديل	تعديل	Match
175.	month-months	شهر	شهرياً-شهور	Match
176.	move	نقل	تنقل	Similar
177.	name	اسم	الاسم	Match
178.	nationality	جنسية	جنسية	Match
179.	necessary	ضروري		Not Match
180.	need	بحاجة إلى		Not Match
181.	network	شبكة الاتصال	شبكة	Match
182.	note	ملحوظة	ملحوظة	Match
183.	notice	تنويه	إشعار	Match
184.	number	رقم	الرقم	Match
185.	objection	اعتراض		Not Match
186.	objectives	الأهداف	الأهداف	Match
187.	obtained	تم الحصول عليها	الحصول	Similar
188.	office	مكتب. مقر. مركز	مكتب. مقر. مركز	Match
189.	operations	عمليات	العمليات	Match
190.	opinion	رأي		Not Match
191.	option	اختيار	اختياري	Similar
192.	organization	منظمة	منظمة	Match
193.	organizer	منظم		Not Match
194.	other	آخر	أخرى	Similar
195.	outstanding	أمتياز		Not Match
196.	page	صفحة	صفحات	Match

197.	paragraph	فقرة	الفقرات	Similar
198.	participant	مشارك	مشارك	Match
199.	participants	المشاركين	مشاركين	Match
200.	pay	دفع		Not Match
201.	performance	أداء	الأداء	Match
202.	period	فترة	المدة	Similar
203.	periodicals	الدوريات	الدوريات	Match
204.	perseverance	مثابرة	مواظبة	Similar
205.	phone	هاتف	الهاتف	Match
206.	place	مكان	مكان	Match
207.	plan	خطة	الخطة	Match
208.	pledge	التعهد	أتعهد	Match
209.	preparation	تجهيز	التجهيزات	Similar
210.	prescribed	المنصوص عليها		Not Match
211.	presentation	عرض	العرض	Match
212.	president	رئيس	رئيس	Match
213.	previous	سابق	السابقة	Similar
214.	print	طباعة	طباعة	Match
215.	procedure-procedures	إجراء	إجراء-إجراءات	Match
216.	process	معالج	عملية	Similar
217.	productivity	إنتاجية	إنتاجية	Match
218.	professor	دكتور جامعي	أستاذ	Match
219.	program	برنامج	برنامج	Match
220.	project	مشروع	مشروع	Match
221.	proportion	نسبة	نسبة	Match

222.	proposals	اقتراحات	اقتراحات	Match
223.	provision	تقديم	تقديم	Match
224.	provisions	أحكام		Not Match
225.	public	عامة	العامة	Match
226.	publication	منشور	المنشورة	Similar
227.	purpose	غرض	لأغراض	Match
228.	purposes	المقاصد	الهدف	Similar
229.	quadrilateral	رباعي		Not Match
230.	quality	جودة	الجودة ـ النوعية	Match
231.	questionnaire	استطلاع	استطلاع	Match
232.	Raise	ربى	رفع	Not Match
233.	rate	معدل	معدل	Match
234.	reasons	أسباب	أسباب	Match
235.	recommendation-recommendations	توصية	توصية-توصيات	Match
236.	reduced	انخفاض	تخفيض	Similar
237.	reference	مرجع	المراجع	Match
238.	registrar	المسجل	مسجل	Match
239.	registration	التسجيل	التسجيل	Match
240.	regulations	قوانين	أنظمة	Similar
241.	related	ذات صلة	المتعلقة	Similar
242.	relevance	ملاءمة	ملائمة	Match
243.	relevant	ذو صلة	علاقة	Similar
244.	report	أبلغ عن	تقرير	Similar
245.	reports	تقارير	تقارير	Match
246.	request	طلب		Not Match

247.	required	مطلوب	المطلوب	Match
248.	research	ابحاث	البحوث	Match
249.	researcher	الباحث	الباحث	Match
250.	resolution	القرار	قرار	Match
251.	resources	موارد	موارد-مصادر	Match
252.	response	استجابة	الاستجابة	Match
253.	responsibility	المسئولية	مسؤولية	Match
254.	result	نتيجة	نتيجة	Match
255.	safety	سلامة	سلامة	Match
256.	scientific	علمي	العلمي	Match
257.	search	بحث	البحث	Match
258.	second	ثانيا		Not Match
259.	section	الجزء	قسم - دائرة	Not Match
260.	sections	الأقسام	الأقسام	Match
261.	selected	المحدد		Not Match
262.	semester	نصف السنة		Not Match
263.	service	الخدمات	خدمات	Match
264.	session	جلسة	الجلسة	Match
265.	set	جلس		Not Match
266.	signature	التوقيع	توقيع	Match
267.	signing	التوقيع	توقيع	Match
268.	site	موقع	موقع	Match
269.	sources	مصادر	مصادر	Match
270.	specialization	تخصص	التخصص	Match
271.	staff	العاملين	العاملين	Match

272.	student	طالب علم	الطالب	Match
273.	studies	دراسات	الدراسات	Match
274.	study	دراسة	دراسة	Match
275.	subject	موضوع	موضوع	Match
276.	successful	ناجح	ناجح	Match
277.	successive	متتالي		Not Match
278.	sufficient	كاف	كافي	Match
279.	suggestions	اقتراحات	اقتراحات	Match
280.	suitable	متكافئ		Not Match
281.	summary	ملخص	ملخص	Match
282.	summer	الصيف	صيفي	Match
283.	supervisor	مشرف	مشرف	Match
284.	support	الدعم	الدعم	Match
285.	system	النظام	نظام	Match
286.	target	استهداف	الهدف	Match
287.	targets	أهداف	أهداف	Match
288.	teaching	تعليم	تعليمية	Similar
289.	teams	فرق		Not Match
290.	technician	فني	الفني	Match
291.	technology	تقنية	تكنولوجيا	Similar
292.	telephone	هاتف	هاتف	Match
293.	terms	شروط	شروط	Match
294.	tests	اختبارات	اختبارات	Match
295.	text	نص		Not Match
296.	time	زمن	الزمن	Match
297.	title	عنوان	عنوان	Match
298.	topic	موضوع	موضوع	Match

299.	total	مجموع	المجموع	Match
300.	training	تدريب	تدريب	Match
301.	transfer	تحويل		Not Match
302.	treated	يعالج		Not Match
303.	trustees	الأمناء	الأمناء	Match
304.	type	اكتب	يكتب	Similar
305.	unit	وحدة	وحدة	Match
306.	university	جامعة	الجامعة	Match
307.	use	استعمال	استعمال	Match
308.	used	مستخدم	مستخدم	Match
309.	value	القيمة	القيمة	Match
310.	version	الإصدار		Not Match
311.	vision	رؤية	الرؤية	Match
312.	vocabulary	مفردات اللغة	مفردات	Match
313.	warning	تحذير	انذار	Match
314.	weak	ضعيف	ضعيف	Match
315.	weakness	ضعف	ضعيف	Similar
316.	whole	كامل	كاملة	Similar
317.	withdrawal	انسحاب	انسحاب	Match
318.	work	عمل	العمل	Match
319.	working	عامل	العاملون	Similar
320.	workplace	مكان العمل	العمل	Match
321.	workshop	ورشة عمل	ورشة	Match
322.	written	مكتوب	مكتوب	Match
323.	year	عام	العام	Match