The Islamic University–Gaza
Research and Postgraduate Affairs
Faculty of Information Technology
Master of Information Technology

الجامعـــــــة الإســـلاميـة – غـزة
شئون البحث العلمي والدراسات العليا
كـليـــة تكنولوجيا المعلومات
ماجستير تكنولوجيا المعلومات

Automatic Linking of Short Arabic Text to Wikipedia Articles

الربط التلقائي للنصوص العربية القصيرة مع مقالات الويكيبيديا

Fatoom M. A. Fayad

Supervised by

Dr. Iyad M. AlAgha

Assistant Professor of Computer Science

A thesis submitted in partial fulfilment
Of the requirements for the degree of
Master of Information Technology

July/2016

Abstract

Given the enormous amount of unstructured texts available on the Web, there has been an emerging need to increase discoverability of and accessibility to these texts. One of the proposed solutions is to annotate texts with information extracted from background knowledge. Wikipedia, the free encyclopedia, has been recently exploited as a background knowledge to dynamically annotate text with complementary information. Given any piece of text the main challenge is how to determine the most relevant information from Wikipedia with the least effort and time. While Wikipedia-based annotation has mainly targeted the English and Latin versions of Wikipedia, little effort has been devoted to annotate Arabic text using the Arabic version of Wikipedia. This work proposes an approach for dynamic linking of Arabic short texts to Wikipedia articles. It reports on the several challenges associated with the design and implementation of the linking approach including the processing and setting up of the Wikipedia's enormous content, the mapping of texts to Wikipedia articles, the problem of article disambiguation, and time efficiency. The proposed approach focuses on short texts because they are generally more difficult to process and annotate than long texts. The proposed approach was assessed over a dataset of 100 short texts gathered from online Arabic articles and then work on it offline. Hyperlinks generated by the approach were compared with the hyperlinks generated by two human raters. The dynamic linking approach achieved **71.79%** accuracy, **74.70%** average precision, and **82.63 %** average recall. A thorough analysis and discussion of the evaluation results are also presented to address the limitations and strengths as well as the recommendations for future improvements. The source code, dataset, and complete experimental results are made available online on: **https://github.com/FatoomMFayad/Dynamic-Linking**

Keywords: *Wikipedia, Arabic, Short texts, Dynamic Linking, Annotation, Anchoring.*

الملخص

في ظل الحجم الهائل من النصوص العربية غير المنظمة المتاحة عبر الإنترنت؛ أصبحت هنالك حاجة متنامية لزيادة قابلية الاكتشاف والوصول لهذه النصوص، ومن أحد الحلول المقترحة هو ربط النصوص بالمعلومات المستخرجة من قواعد المعرفة، وقد تم استخدام ويكيبيديا (الموسوعة الحرة) كقاعدة معرفة للربط التلقائي للنصوص بالمعلومات الإضافية مؤخراً. ومن أهم التحديات عند التعامل مع أي نص هو كيفية تحديد المعلومات التوضيحية الأكثر صلة بهذا النص من الويكيبيديا بأقل وقت وجهد. وعلى الرغم من وجود جهود كثيرة متعلقة بالربط القائم على الويكيبيديا إلا إنها استهدفت بشكل أساسي النسخ الإنجليزية واللاتينية وكان هنالك جهد أقل متعلق بالنسخة العربية من ويكيبيديا. يهدف هذا العمل إلى اقتراح حل للربط التلقائي للنصوص العربية القصيرة بمقالات الويكيبيديا ويتطرق إلى التحديات المتعلقة بتطوير وتنفيذ عملية الربط ومن ضمنها تجهيز الويكيبيديا والتعامل مع محتواها الضخم، والربط مع مقالاتها وتوضيح الغموض المتعلق بمعنى المقالات وكذلك الوقت المستغرق في عملية الربط. يركز العمل المقترح أيضاً على النصوص القصيرة كونها أكثر صعوبة من النصوص الطويلة من ناحية المعالجة والربط.

تم تقييم النظام المقترح باستخدام مجموعة من 100 نص عربي قصير والتي تم تجميعها من عدة مواقع إلكترونية عربية من خلال مقارنة الروابط الناتجة من النظام مع الروابط المقترحة من قبل إثنين من المقيمين وقد حقق النظام نسبة دقة 71.79% .74.70% precision و82.63% recall. وبعد الحصول على النتائج تم تحليلها ومناقشتها بشكل متعمق من أجل توضيح نقاط القصور والقوة والتوصيات المقترحة من أجل التطوير المستقبلي على العمل.

من أجل الحصول على الكود، مجموعة النصوص ونتائج التجارب يمكن الرجوع للموقع:

https://github.com/FatoomMFayad/Dynamic-Linking

كلمات مفتاحية: ويكيبيديا، عربية، نصوص قصيرة، ربط، تلقائي، توضيح، ربط.

Dedication

To the sake of ALLAH, my creator and my Lord.

To My teacher, leader, intercessor, and beloved prophet Muhammed (peace and blessings be upon him).

To my beloved family: parents, sisters, and brother, whose affection, love, encouragement and prays of day and night make me able to get such success and honor.

To my gorgeous sister Hanan, my shining source of hope and goodness.

To my nieces and nephews, with hope for a bright future.

To my magnificent friend Inas, for her support, care, and endless love.

To my friends, who encourage and support me.

Acknowledgment

At the very outset, all my prayers and thankfulness are to Allah the almighty for facilitating this work and for granting me the opportunity to be surrounded by great and helpful people.

I would like to express my everlasting gratitude to my supervisor, **Dr. Iyad Mohammed Al Agha** for his valuable encouragement, guidance and monitoring, I would like to thank him for pushing me to the limits on this thesis, for spending valuable time on my drafts and for ceaselessly providing me with feedback and discussion. Without which this work would not have reached the point of fruition, so I ask Allah to reward him on my behalf.

To my father, who encouraged me to be the best I can be, to have high expectations and to fight hard for what I believe. The man to whom I will be grateful, for the rest of my life.

To my mother; my source of inspiration for the whole of my life. I would never forget her continuous prayer for the sake of my success.

A special thanks to my family. Words cannot express how grateful I am to my sisters, Tagreed, Takween, Eman, Bassma, and Hanan, to my brother Ahmed, and sister-in-law Kholoud .Thank you for supporting me for everything, and especially I can't thank you enough for encouraging me throughout this experience.

To my beloved friend Inas, I would like to express my thanks for being such a great friend.

To my nieces and nephews, Saja, Hala, Mahmoud, Hamza, Batoul, Mays, Mohy, Ayham, and Omar for being such a great kids always cheering me up.

To my friend Eman S. Fayyad for being always supportive, and for helping me to complete this work.

To my friend and spiritual mother Rawya, thanks for being tender and supportive.

To my manager and my programming godfather, Eng. Marwan Abu Shighiba, and my colleagues in Computer Center Palestine Technical College- Deir Elbalah. Your advice on my career has been invaluable and helpful.

Finally, my appreciation goes to all colleagues and friends for their suggestions and encouragement during the writing of this thesis.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

API	Application Programming Interface
ENB	Extended Naïve Bayes
ER	Entity Ranking
IPFG	Interest Prediction Factor Graph
IR	Information Retrieval
JWPL	Java Wikipedia Library
NLP	Natural Language Processing
OWL	Web Ontology Language
RDF	Resource Description Framework
TF-IDF	Text Frequency Inverse Document Frequency
URL	Uniform Resource Locator
XML	Extensible Markup Language

Chapter 1
Introduction

Chapter 1
Introduction

1.1 Introduction

This research considers the problem of interlinking unstructured Arabic text to knowledge resources such as Arabic Wikipedia. Linking, or sometimes referred as annotation or anchoring is the process of linking terms in documents to their corresponding resources from external knowledge bases (Ramudu & Murty, 2012). For example, terms in Web documents can be converted to hyperlinks leading to their corresponding Wikipedia articles, so that readers can access complementary information related to the text being read. Linking terms to external knowledge resources provide several benefits such as increased discoverability and accessibility, and hence the utility of information. Furthermore, it can be useful to add semantics to documents and hence allow machines to process documents in an intelligent manner (Sriharee, 2014)

Considering the numerous side of the Web, the manual linking of documents on the Web can be very tedious and time consuming. For each document, authors or web developers should find out relevant resources, e.g. other Web articles that contain proper definitions or explanations for the key terms. These difficulties have motivated several researchers to explore approaches for dynamic hyperlinking of text. These approaches have primarily focused on mining the Web for proper information to link with terms in the text (Adnan, Warren, & Orr, 2013; Navigli & Ponzetto, 2012)

With the advent of massive Web-based knowledge bases and encyclopedias such as Wikipedia, DBpedia and Linked Open Data, it has been possible to obtain comprehensive details on almost every topics. These knowledge bases have been used by several research efforts to facilitate dynamic linking of document (Barrena, Donostia, Soroa, & Agirre, 2015; Geiß, Spitz, & Gertz, 2015; Mirizzi, Ragone, Di Noia, & Di Sciascio, 2010; Moro, Raganato, & Navigli, 2014). Given any piece of text the main challenge is how to determine the most relevant explanatory information from knowledge bases with the least effort and time. Existing research investigating

dynamic linking has explored the use of different knowledge bases such as DBpedia (Mirizzi et al., 2010), Wikipedia (Ikikat, Gurhan, & Diri, 2015), WordNet (Passant, 2010) and ontologies (Sriharee, 2014). Several techniques have been also used to map words in documents to their matching resources/articles/concepts from knowledge bases. However, the majority of these efforts have focused on English or Latin-based languages. Little efforts; however, have been done to explore the dynamic annotation of Arabic text to external knowledge resource. The limited efforts in this field can be attributed to: First, the limited number of knowledge bases that are publically-available in Arabic. Although the Arabic version of Wikipedia and DBpedia have become available and are rapidly growing, the sizes of these knowledge bases are still largely small when compared to versions of other languages such as English or French. In addition, the number of domain ontologies that are expressed in Arabic is very limited as compared to the ontologies expressed in English. Second, the lack of accurate and efficient Natural Language Processing (***NLP***) tools for Arabic, as compared to the tools available for English, has notably slowed down research in Arabic NLP in general (Farghaly & Shaalan, 2009).To our knowledge, Arabic Wikipedia is the largest Arabic knowledge source that covers most of the technical and non-technical topics, events that have happened, and topics related to most of the domain areas. Despite its small size as compared to other Wikipedia versions, Arabic Wikipedia is still the best choice for annotating Arabic text. This is due to its high coverage of many domains. It also supports easy extraction of lexical information due to its relatively high structured content.

This research aims to build a dynamic linking approach for short Arabic texts by exploiting Arabic Wikipedia as background knowledge. Given any short Arabic text the proposed approach searches Wikipedia for the articles that best describe the key terms within the text. It also tries to handle the various challenges associated with the linking process including the processing of the Wikipedia's massive content, the mapping to Wikipedia articles, the ambiguity of terms and the time efficiency. It focuses on short texts because they are generally more difficult to process and annotate than long texts. This is because short texts often do not provide adequate information that enables the application of statistical or machine learning techniques which have been extensively used for the annotation of long texts. For example, it is difficult to

apply frequency-based techniques such as TF-IDF to identify keywords in short text. Therefore, the proposed approach will employ alternative techniques to extract keywords from short text and map them to Wikipedia articles. Furthermore, by considering short texts, the proposed approach aims to apply the linking approach on social media content which mainly consists of short tweets and posts. The proposed approach can be also applied on long text by; for example, using a sliding window of predefined length over the long text.

The source code of the proposed approach and the installation instructions are available online on **(https://github.com/FatoomMFayad/Dynamic-Linking)** and are free to use for research and academic purposes. To our knowledge, this is the first effort that aims to offer dynamic linking of Arabic text to Wikipedia articles. While the English version of Wikipedia has been widely utilized in several research areas related to information retrieval, there has been little efforts to utilize the Arabic version of Wikipedia for similar research on Arabic text. The proposed approach is expected to act as a baseline for the research tackling Wikipedia-based annotation of Arabic text.

It was assessed over a dataset of 100 short texts gathered from online Arabic articles like Aljazeera Channel, TRTalarabiya, RTarabic, and Sky News Arabia. Hyperlinks generated by the approach was compared with the hyperlinks generated by human subjects. Our linking approach achieved (**71.79%**) accuracy, (**74.70%**) average precision, and (**82.63 %**) average recall; while the error rate was (**28.21 %**).

1.2 Statement of the problem

The main problem addressed in this research is how to exploit Arabic Wikipedia to offer dynamic annotation or linking of short Arabic text. While many approaches have been proposed to annotate English text to knowledge bases such as Wikipedia, DBpedia and ontologies, no effort, to our knowledge, has explored the use of Arabic version of Wikipedia for dynamic linking of Arabic text.

1.3 Objectives

In this section, we present both main and specific objectives of the research work.

1.3.1 Main objective

The main goal of this research is to design and implement a dynamic linking approach to annotate Arabic short-texts with pertinent Wikipedia articles.

1.3.2 Specific objectives

1. Explore ways to process and access the large content of Arabic Wikipedia in an efficient and accurate manner.
2. Study the existing approaches used for linking text to external knowledge bases, and build on them to use the Arabic version of Wikipedia.
3. Provide an approach to map Arabic terms in text to the most relevant articles from Wikipedia.
4. Address the unique characteristics of short text, such as the inability to apply statistical or machine learning techniques.

1.4 Importance of Research

1. Connect the non-structured short Arabic texts to the content of Wikipedia articles, so that this data can be reused and shared across community boundaries, applications, and enterprises as a Linked Data.

2. Offer an extended information for the normal user of social media and search engines by linking the short texts with related Wikipedia articles.

1.5 Scope and limitations

1. This work utilizes only the Arabic Wikipedia.

2. To provide efficient access to Wikipedia content, we used Wikipedia XML dump that is stored in a local relational database.

3. Our approach considered linking of Arabic short texts.

4. Our approach considered only the standard Arabic language, non-standard Arabic language is out of scope for this research.

5. The evaluation of the linking approach was done using a specific dataset gathered from Arabic news pages on twitter such as Aljazeera, Alarabiya, and Sky News Arabia. It was not possible to conduct a comparative study due to the lack of similar linking approaches of Arabic text.

1.6 Research Contributions

The work in this thesis has the following research contributions:

1- This is the first work, to our knowledge, that explores the annotation of Arabic text with links to Wikipedia content. Arabic Wikipedia has been exploited only recently by the Arab computer researchers and few efforts from the literature have tried to interface to the Arabic version of Wikipedia for different purposes, such as determining relations between topics (Kanan et al., 2015), named entity recognition (Althobaiti, Kruschwitz, & Poesio, 2014) and ontology generation (Al-Rajebah & Al-Khalifa, 2014).

2- The work builds on the techniques used in counterpart efforts that tackled the English and Latin versions of Wikipedia and tried to adapt them to the Arabic version of Wikipedia.

3- It proposes an in-depth evaluation of our linking approach and explored the potential shortcomings and strengths. This detailed evaluation can inform Arab research community with the various design options, challenges and recommendations when designing similar approaches. The evaluation also provides insight into several issues not addressed in existing efforts such as issues related to article disambiguation and filtering.

4- Present a novel approach that we can consider it as a guideline for the future efforts in utilizing Arabic Wikipedia structure in real life applications.

5- Generate a standard dataset of Arabic short-texts.

6- Source code for the developed approach is available online for researchers.

1.7 Structure of Thesis

The thesis consists of five chapters. The chapters are organized in general as follows:

Chapter 1: Introduction: This chapter considers interlinking unstructed Arabic text to knowledge sources such as Arabic Wikipedia. It also discusses the difficulties of using Arabic Wikipedia due to the complex problems related to the speciality of Arabic language.

Chapter 2: Literature Survey: This chapter focuses on the background and the concepts of Wikipedia. As well as the related work, that used Wikipedia as a knowledge source.

Chapter 3: Methodology: This chapter presents the developed linking approach by describing the steps of linking process. And explains a real scenario of using the approach and the results of each phase.

Chapter 4: Results and Discussion: this chapter explains the gathered data set of Arabic short texts, explains the used evaluation metrics, and discusses the results focusing on the sources of errors.

Chapter 5: Conclusions: Discusses the conclusions and presents possible future works.

Chapter 2
Literature Review

Chapter 2
Literature Review

2.1 State of the art

The world-wide-web has become the largest ever free-access information repository with billions of web pages (Abdeen & Tolba, 2010). Moreover the challenging thing is that the Internet does not stay static and it always develops dynamically and rapidly(Badie, Azimzadeh, Zahedi, & Samuri, 2014). Today's numerous data repository needs novel approaches to mine such data efficiently and effectively, recently many efforts attempted to go beyond this paradigm to improve the mining and indexing of textual data (Ferragina & Scaiella, 2010)**.** Researchers have concentrated on how to provide a better accuracy and quicker retrieval based on keyword and textual semantics(Chan, Baciu, & Mak, 2009)**.**Many systems today espouse Wikipedia pages (or derivatives of them) as topics and do the annotation process by hyperlinking the selected sequences of terms with Wikipedia pages that are related to their meaning(Ferragina & Scaiella, 2010).Wikipedia is one of the largest knowledge source. It covers most of the technical and non-technical topics, events that have happened, topics related to most of the domain areas(Ramudu & Murty, 2012).Many researches targeted Wikipedia as a knowledge source in many applications especially in ***NLP***, information retrieval (***IR***), semantic annotation systems, and so on. There are many reasons for using Wikipedia:

1. It is publicly available to access and update in many languages, including Arabic.

2. Easy extraction of lexical information from Wikipedia due to its relatively high structured content.

3. High coverage for many domains, including medicine, News, Sports, etc.

4. The knowledge contributions progress in any domain will be reflected immediately in Wikipedia knowledge(Ramudu & Murty, 2012; Vivaldi & Rodriguez, 2014).

One of the most important reasons for using Wikipedia as knowledge source is its well-structured hierarchy. This hierarchy consists of main categories that can be divided to sub-categories, each one of the article titles(concepts)can be classified under one or more category, redirects are used to group all the equivalent terms for a given concept(Ramudu & Murty, 2012).Wikipedia contains 271 different languages. For example, the Arabic Wikipedia ranked 27th with more than 771,000 articles (Wikimedia, 2015). This produced the motivation to employ Arabic Wikipedia's underlying structure in many applications such as annotation of short texts like search engine results, social media tweets and posts, news, and so on. To this end, it is useful to overview the most important aspects of the Wikipedia structure(Alotaibi & Lee, 2013).

Wikipedia structure:

1. **Articles**: can be one of the following:

    1. **Normal articles:** Each article has a unique title and consists of authentic content; i.e. textual data, tables, images, items and links which related to the concept (article's title).

    2. **Redirected articles:** These contain a particular tag to redirect the inquirer to a normal article. For the redirect article titled (**بريطانيا العظمى**/ Great Britain), there is a redirected tag to (**المملكة المتحدة,** United Kingdom). This tag is written as #REDIRECTED [[**المملكة المتحدة**]].

    3. **Disambiguation articles:** List all article titles that share ambiguities.

2. **Links types**: There are two types of Wikipedia's links:

    1. **Non-piped links:** This type of links indicates that the display phrase of the link and the article's title are the same. For example, [[Paris]].

    2. **Piped links:** This type of links denotes that the text that appears in the contextual data is different from the article it refers. For example, [[USA]] [[United States of America]] where "USA" appears in the

display text, when “United States of America” refers to the article’s title.

3. **Connectivity:** Used links, in the contextual data of any Wikipedia article, provide connectivity, and thereby the Wikipedia’s structure that will be utilized in the research.

4. **Article Features:** Here we specify three features of Wikipedia articles(Technology, 2014):

    1. **Outlinks:** For a Wikipedia article **w**, the outlinks of w is the set of articles $\mathbf{O_w}$ that w links to by hyperlinks.

    2. **Inlinks:** For a Wikipedia article **w**, the inlinks of w is the set of articles $\mathbf{I_w}$ that link to w by hyperlinks.

    3. **Categories:** For a Wikipedia article **w**, the categories of w is the set of categories $\mathbf{C_w}$ that w belongs to.

Social media is a user formulated data such as Facebook posts, tweets, blogs, and so on. Social media data has now become widely used and because of its importance in many information retrieval applications, it has encountered significant attention which has motivated many researchers to employ such data in scientific researches(Kaplan & Haenlein, 2010). We employed Arabic Wikipedia with its growing and structured data to design and implement a dynamic linking service to annotate Arabic short-texts gathered from social media with pertinent Wikipedia pages. We will offer the proposed annotation service to the Arab research community through an online web service that researchers can utilize for research purposes.

2.2 Review of related works

It is usually worth considering the efforts of other researchers and checking if we can refine and extend the existing knowledge of our particular domain. There are many researches examine the use of Wikipedia as knowledge source in information retrieval, natural language processing, and semantic annotation, or combine techniques of preceding fields in a given domain. Many efforts like (Demartini, Firan, Iofciu,

Krestel, & Nejdl, 2008; Gruetze, Kasneci, Zuo, & Naumann, 2014; Zhang, Zhang, Wu, Huang, & Ma, 2014) targeted the English version of Wikipedia due to its huge and rich structured content, while fewer works (Al-Rajebah, Al-Khalifa, & Al-Salman, 2011; Shams Eldin & El-Beltagy, 2013) employed the Arabic Wikipedia in scientific research. As a result, it stills a hot topic in the Arabic domain.

2.2.1 Using Wikipedia in Computer Science

While (Demartini et al., 2008) presented a formal model to determine entities as well as a complete ***Entity Ranking*** (***ER***) system. It also provided a set of algorithms based on the proposed system and assessed their retrieval effectiveness using an existing environment that made an automatic evaluation environment for ER based on Wikipedia. Another research (Gruetze et al., 2014) addressed the problem of clustering search results of ambiguous person-name queries by proposing a framework build on the idea that Wikipedia pages about persons, can be bootstrapped to deal with the above problem as a classifaction problem where results are mapped to the most relevant profile. The researchers (Zhang et al., 2014) presented an Interest Prediction Factor Graph (IPFG) model, which is identified by the editor's social properities, hyperlinks between Wikipedia entitiesm categories of an entry and other important features. The parameters of the model are learned using gradient descent algorithm and Loopy Sum-Product algorithm.The researchers analyzed the impact of different number of editor's properities on the prediction accuracy of the proposed model. The evaluation results showed that the proposed model outperform a collaborative recommendation algoithm on accuracy in terms of F1-measure. An earlier effort (Ramudu & Murty, 2012) proposed a topic based classifiaction of academic researchers using Wikipedia.The researchers first generated a corpus for academic researchers because there was no existing standard corpus.Then they created a complete academic topics dictionary.After that they scanned these topics in researchers' publications to create document-topic mapping vector.Then soft clustering algorithm is used for grouping related researcher's. From the resulted clusters brief topics for each cluster were gained and the researchers were linked if they belong to the same cluster, through the common topics. The evaluation showed that the developed approach was useful in retieving profile details of resaercher's

having expertise in a particular topic. And the research (Vivaldi & Rodriguez, 2014) proposed an improved method for obtaining terms from potentially any domain using the Wikipedia graph as a knowledge source. The researches applied their approach in the medical domain and Arabic language. The evaluation results showed that the coverage and precision of the obtained terms are satisfactory. The effort in (Cosh, 2013) served the traveling and tourism field by presenting an application of NLP algorithms that specifies the main features of each destination automatically and then recommend other destinations that share similar features using Wikipedia articles. First, a list of stop words are tremmed and many word stemming techniques are investigaetd to speed up the processing. Second, each destination (or article) is represented as a set of keywords with corresponding log likelihood values. The application was assessed by user driven comaprisons to improve the results by allowing human interaction, so the author was satisfied by the results.

2.2.2 Using Arabic Wikipedia

The researchers (Al-Rajebah et al., 2011) proposed a system that employed Arabic Wikipedia to build an ontological model that represnts concepts and their semantic relations for a specific domain. The system architecture consists of XML Parser to parse Wikipedia XML dump file downloaded from internet to extract for each article a list of features that are smenaticlly related to.In the proposed system, those features are extracted from the list of categories an article belongs to and the infoboxes supplied by Wikipedia. The genrated features and semantic unit are stored as a graph and then given to ontology generator to convert it to the appropriate OWL format. The system was evaluated using two measures: human judges and precision and the results proved its goodness as it yielded suitable results.

2.2.3 Text tagging

2.2.3.1 Text tagging using ontologies

The work in (Sriharee, 2014) was an auto-tagging system. The system recommends ontological tags for articles, the tagging process contains:

- Classification of the given article to the most relevant domain using cosine similarity.

- Moreover, tag selection process: first matched the extracted tag of the article with the concepts specified in tag ontology of relevant domain. Then the suggestion of tags based on their significance.

The evaluation of the system used 140 articles in Thai language, and compared the results with the result of manual tagging. The average recall of the experiment was 0.98 and precision was 0.85, and the classification's accuracy was 90%.

2.2.3.2 Text tagging using Wikipedia

While (Ikikat et al., 2015) presented an application (Detective – Dedektif) for deciding the Wikipedia word and phrases which will be related with other Turkish or English Wikipedia articles, and generating a hyperlinks automatically. The developed application took a Turkish Wikipedia pages as inputs and processed it in two steps: the first step finding the roots of words in Turkish, then the words suitable for keyword extraction are translated to English. In the second step, English Semantic Similarity was used to define the semantic similarity between translations of the Wikipedia article title and the selected word or phrase. After obtaining the semantic similaity measure, the words which are more than a given threshold value were researched on Wikipedia and links were created inside the text. The evaluation results showed that the success of the Turkish version is better than English version.

Another effort (Scheau, Rebedea, Chiru, & Trausan-Matu, 2010) provided a method to extract semantic relations between the names of the Wikipedia's articles. Then it employed these relations in defining the rank of Google search engine's results. The proposed method divided into two main phases: a preprocessing phase using TF-IDF in order to obtain semantic similarity between concepts based on the redirect pages and the internal links. A second phase for processing the search engine's results taking into account the semantic relations, which obtained in the previous phase and the user query. The proposed system's performance was measured by the satisfaction of the user uses both systems (Google, and the proposed system) and compare the results.

The researchers in (Zhiqiang, Werimin, & Zhenhua, 2009) provided a web-based method for measuring the semantic similarity between words. The presented method employed the Wikipedia snippets to calculate the semantic similarity between words by utilizing: preprocessing techniques, such as stemmer algorithms and stop words, Cosine Similarity, and Term-Frequency Inverse Document TF-IDF. The method assessed by using Rubenstein-Goodenough benchmark dataset. The results showed that method improves the correlation and accuracy for measuring the semantic similarity between words.

2.2.3.3 Text tagging using LinkedData

The research effort in (Mirizzi et al., 2010) provided a framework for annotating and retrieving web content by using semantic tagging with DBpedia. The framework consists of:

- A tool that can suggest similar tags to the users.
- A hybrid ranker of DBpedia resources relying on RDF graph and results of search engines like Google, Yahoo, and Bing.
- A relative ranking system that using the similarity relation between resources.

The main steps of the proposed approach are the following: 1) mapping of keywords to corresponding DBpedia resources. 2) After this mapping, the researchers can associate a defined semantics to keywords so that they can enrich the "meaning" of the keywords by utilizing the DBpedia ontological structure.

The system compared to four different algorithms; 50 users evaluated the five algorithms. The results showed that the proposed system performed better than the remaining algorithms.

2.2.3.4 Tagging of Arabic Texts

While (Shams Eldin & El-Beltagy, 2013) used Wikipedia pages names, redirects, and properties to build a concept dictionary or pool of meaningful tags,

properties, and re-directs to build a pool of meaningful tags. Complex and efficient matching techniques were then utilized to find out text segmants in news stories that match entries in the resulted tag pool. The main idea behind the work is that every Wikipedia article maps to real life concept or entity. The proposed approach consists of ywo main phases: 1) Using Wikipedia to build a pool of meaningful tags by first, extract the content of Wikipedia (inlinks, outlinks, and titles). Second, discard the non-important Wikipedia entries to spped up the tagging process. 2) Tagging input text: the goal of this phase is to match fragments of input text to tags in the Wikipedia constructed dictionary. The main steps of the automatic tagging are: sentence tokenization, extracting candidate anchors, and finally matching candidate anchors with the constrcuted dictionary. The system was assessed by comparing the tags that it genrates to those generated by MSN Arabic news. The results showed that the proposed method outperform MSN Arabic news. Another similar effort (Zaraket & Jaber, 2013) presented a morphological tagging for Arabic texts. The researchers developed a Graphical User Interface tool that allows the user to make a manual and automatic tagging, can edit the resulted tag, and calculate accuracy of the results.

2.2.3.5 Tagging Short Texts

While (Ferragina & Scaiella, 2012) presented a system that can enhance a plaintext with related hyperlinks to Wikipedia pages. This system can annotate English and Italian short texts (such as snippets of search engine results, tweets, news, and so on) on the fly using bag-of-words paradigm. The linking of annotated topics taking advantage of relatedness measures defined on the Wikipedia structure. The proposed approach basically made use of the inlinks and outlinks of the Wikipedia articles to measure the semantic relatedness. The higher the links between two articles the higher the semantic relatedness between them.The result is a sophisticated semantic graph of topics. The system compared to known datasets and the results showed that it could outperform the known by 8 percent in F-measure test. The research work in (Derczynski, Ritter, Clark, & Bontcheva, 2013) compared many of English part-of-speech existing taggers against Twitter data. Then proposed a data-intensive approach based on noticed difficulties with microblog tagging like word sparsity. The researchers claimed that their tagger reduces the error rate at token and sentence levels.

Another research (Nakamura, Shirakawa, Hara, & Nishio, 2014) provided two methods SimpleMap and ProbMapto that measure the semantic similarity for multilingual short texts by intergating inter-language links of Wikipedia into extended naïve Bayes (ENB). First the proposed approach united the language space of input text into a single language space which was in this research English, which expanding the semantic information for short texts and uniting the language space of the given multilingual texts. The proposed two methods finally generated a vector of related English entites to the input text as its semantic representation. This vector can be utilized to measure the semantic similiarity between multilingual short texts using any known metric like cosine or tf-idf.

2.2.3.6 Linking Social Media to Wikipedia

The effort in (Ren et al., 2013) presented a forensic text analysis approach that utilized entity linking to specify social media entities and link them to Wikipedia articles. The developed approach consists of two phases; in the first phase, the researchers developed a state-of-the-art semantic search approaches that link entity in micro-blog posts to related Wikipedia articles, while in the second phase performed a personalized time-aware tweets summarization. This summarization connecting entities to Wikipedia articles based on Entity Linking Task, this connection used to enhance the tweets with the most important sentences of the connected Wikipedia articles. The two steps of the proposed approach evaluated separately. First, the semantic linking step assessed by measuring the $B^{\wedge 3}$ and F1 measures of Text Analysis Conference Knowledge Base Population evaluation campaign, and the results showed that the followed clustering technique improves the overall performance of the approach. Then the personalized tweets summarization step assessed by using a twitter dataset that contains tweets and social relations. And the results showed that proposed approach outperforms the other baselines in different number of selected tweets per period.

Based on the above discussion of previous works, we believe that no efforts have explored the use of Wikipedia as a knowledge knowledge source to annotate Arabic short texts to enrich the language specific knowledge and make it available for usage as Linked Data. While our approach builds on existing efforts that worked on English

Wikipedia, we aim to address the use of Arabic Wikipedia and build the infrastructure that enables for fast access and search in Arabic Wikipedia. We prefer to use Arabic Wikipedia rather than using DBpedia because the Arabic version of Wikipedia covers more concepts and offers a well structured content that enables interlinking existing web contents. While Arabic DBpedia stills mature and needs more efforts to make it available for usage as Linked Data instead of relying on English DBpedia only as a knowledge source(Bahanshal & Al-Khalifa, 2013).

Chapter 3
Methodology

Chapter 3

Methodology

3.1 Introduction

This chapter presents the approach for dynamic linking of Arabic short texts by using Wikipedia. It reports the detailed steps of the linking process which include: First, the configuration and setting up of Arabic Wikipedia and the indexing of the content. Second, the preprocessing of the queried short text. Third, mapping the short text to Wikipedia content. Fourth, ambiguity resolution and filtering. . It also presents a full usage scenario of the approach.

3.2 Setting up and Configuring Arabic Wikipedia

This section briefly explains the configuration needed for our linking approach. This configuration includes the preprocessing of Arabic Wikipedia to enable for fast information access and retrieval. Note that all the configuration settings are performed only once.

3.2.1 Parsing and information extraction from Arabic Wikipedia XML Dump

This sections briefly explains the configuration we undertook to enable for fast access and search over Wikipedia content during the linking process. Note that the configuration steps should be performed only once**.** To deal with Arabic Wikipedia and extract information from it, we cannot rely on the online version of Wikipedia as this will be time consuming. Therefore, we opted to work offline by downloading the XML dump of Arabic Wikipedia (05 March 2016 version). The information about the downloaded dump is shown in Table 3.1.

Table 3.1: Information about the Downloaded Dump

XML Dump File Size	**1.57 GB**
Size after extraction	**22 GB**
Number of Pages	**869453**
Number of Categories	**164497**
Number of Categories-inlinks	**506605**
Number of Categories-outlinks	**506605**
Number of Page-inlinks	**55556636**
Number of Page-outlinks	**55556636**
Number of Category-pages	**4317315**
Number of Page-redirects	**9214**

The XML dump file was then parsed so that relevant information were extracted and stored in a local database. The database contains tables for pages, categories, category-inlinks, category-outlinks, category-pages, metadata, page-inlinks, page-outlinks, page-redirects, and page-mapline. Information will be then extracted by querying the database.***JWPL (Java Wikipedia Library)*** was used to parse and extract information from the XML dump file. ***JWPL*** is an open source Java-based application programming interface that offers free access to all Wikipedia available information. In our research we utilize only JWPL Core **(API, and DateMachine)** (Zesch, Müller, & Gurevych, 2008)**.**

3.2.2 Indexing of Wikipedia content

The core step of the dynamic linking approach is the mapping process which should match the Arabic words with relevant articles from Wikipedia. The mapping process; however, should be performed rapidly without incurring significant time delay.

To enable for fast access and search over the Wikipedia content, we indexed the Wikipedia articles titles, articles inlinks, and articles outlinks by using the Apache Lucene search engine. We used the StandardAnalyzer which supports general token types but does not support word stemming as an analyzer for indexing the above contents. The result of the indexing process is file systems stored on the local machine hard drive. While the Lucene Search API deals with the local database once in indexing process to produce the file systems, any future search query will use the indexed files to return the results.

Apache Lucene (Lucene, 2016) is an open source java-based search engine library. It offers a high-performance cross-platform full-text solution. Content of Wikipedia articles were stored in the local database, but their indices, generated by Lucene, were stored in the file systems. Both Wikipedia Articles contents and the file systems indices put togother in Wiki DB as shown in Figure 3.2.

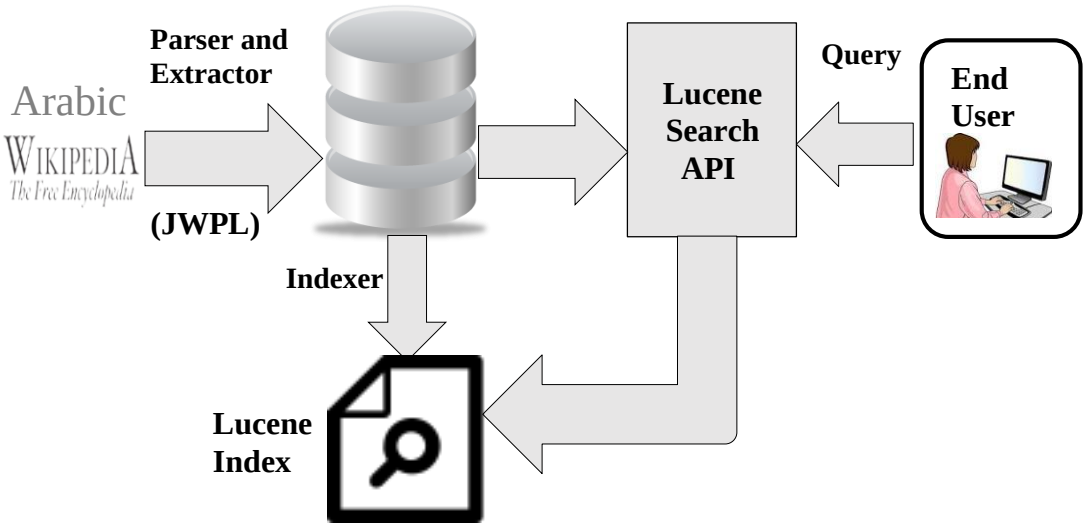

Figure3.1 : Setting up Arabic Wikipedia

3.3 Dynamic Linking of Arabic Short Texts

The dynamic linking approach is illustrated in Figure 3.2 explained as the following:

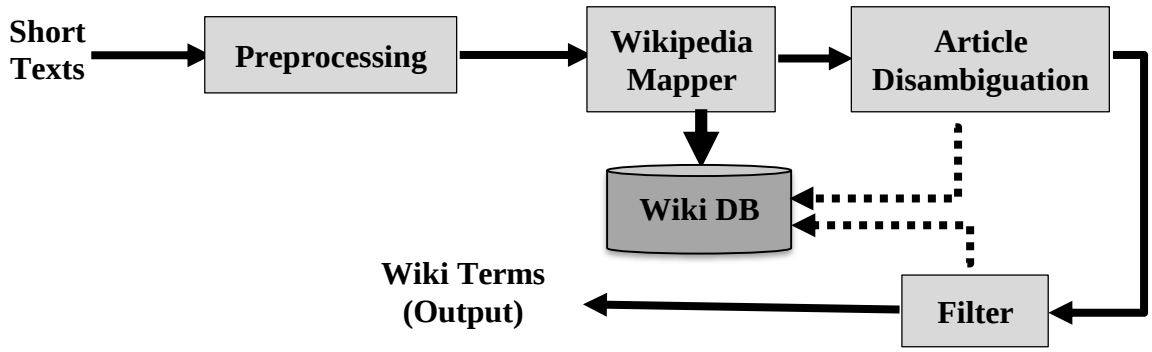

Figure 3.2 : Dynamic Linking Process

3.3.1 Text Preprocessing

Preprocessing is an important task in information retrieving (IR); the main purpose of text preprocessing is to obtain the main features or key terms from text. We applied two preprocessing steps on the input Arabic short text:

- **Stop-word removal**

Stop words are commonly used words that are frequently appeared in a corpus. Such words increase the size of the text and removing them doesn't affect

the retrieving efficiency(Al-Shalabi, Kanaan, Jaam, Hasnah, & Hilat, 2004). We applied a stop-word removal algorithm to reduce the size of the corpus and improve the retrieving efficiency. First, the algorithm read a file containing a list of Arabic stop-words, numbers, less than two-characters words. After that, it iterates over the input text and remove all the listed words. For example, the input text **"ريال مدريد يفوز على برشلونة في الكلاسيكو" "Real Madrid beats Barcelona in Classico"**. The output text will be **"ريال مدريد يفوز برشلونة الكلاسيكو"** after removing **"على" "on"** and **"في" "in".**

- **Text Tokenization, Normalization and N-grams Generation**

After removing stop words, the text was tokenized and normalized. Normalization aims to unify the Arabic letters that appear in different formats (e.g. replacing "أ"with "ا"and "ه"with "ة"). This will enable for better matching with the Wikipedia articles.

To help map phrases in the input text, the text was then split into a set of grams. Grams are phrases consisting of two or more subsequent words from the input text. The aim of generating grams is to increase the matching rate with Wikipedia content by generating all possible combinations of words and map them to Wikipedia articles. For simplicity, we limit the length of the generated grams to be equal to or less than three**.** For example, if the input text is **" ريال مدريد يفوز على برشلونة في الكلاسيكو",** the resulting grams will be **ريال مدريد يفوز- مدريد يفوز برشلونة- يفوز برشلونة الكلاسيكو – ريال مدريد – مدريد يفوز – يفوز برشلونة - برشلونة الكلاسيكو- ريال – مدريد – يفوز – برشلونة – الكلاسيكو.**

3.3.2 Wikipedia Mapper

The Wikipedia mapper (see Figure 3.2) is responsible for mapping n-grams generated from the preprocessing phase to Wikipedia articles. The aim is to identify the Wikipedia articles from which the links should made. Candidate n-grams will be later converted to links to the corresponding Wikipedia articles. For each gram from the input text, the mapper retrieves all Wikipedia articles whose title contain the gram. The mapping with Wikipedia content starts from the highest n-grams. The assumption

here is that longer phrases will represent more specific descriptors than shorter ones. Then candidate grams that are substrings of other longer n-grams are ignored.
For example, from the phrases resulted from the previous phase, for the detected term **"الكلاسيكو",** the word **"الكلاسيكو"** is chosen and **الكلاسيكو يفوز برشلونة andبرشلونة الكلاسيكو** are neglected because they have no articles related to them. The following table shows the word **"الكلاسيكو"** with all possible articles related to it. Where each article has a unique id number.

Table 3.2: Term Detection Example

Term	**Article ID**	**Article Title**
الكلاسيكو	192553	الكلاسيكو
	2319588	الكلاسيكو الألماني
	2778204	الكلاسيكو المغربي

- **Article Disambiguation**

When matching phrases to Wikipedia articles, it is likely that a single phrase matches with multiple Wikipedia articles. Furthermore, several Wikipedia articles may be retrieved when their titles contain the same phrase. In fact, the Wikipedia Mapper may match a single phrase to a large number of Wikipedia articles. For example, the phrase **"ريال مدريد" "Real Madrid"**was matched with 43 Wikipedia article which all contain the phrase **"ريال مدريد"** in their titles. Some of these articles are: **بطولات ريال مدريد، قائمة مدربي ريال مدريد، نادي ريال مدريد، لاعبي ريال مدريد، جماهير نادي ريال مدريد، تاريخ ريال مدريد، قناة ريال مدريد.** In addition, a term from input text may be ambiguous in the sense that it has multiple meanings. Such an ambiguous term may be mapped to multiple Wikipedia pages, each of which denotes a different meaning. For example, the word **"طرابلس"** matches with at least two Wikipedia articles, one denotes the Lebanese city while the other denotes the capital of Libya. In dynamic linking of text; however, a phrase should be linked to a single Wikipedia article. Thus, it is necessary to assure that the detected

phrases are linked to the most relevant articles among all probable articles related to this term in Wikipedia.

When mapping terms in text to Wikipedia links, our assumption is that there should be a collective agreement or relatedness among the detected Wikipedia links. If a term is mistakenly mapped to an invalid article, this article is likely to have a low relatedness with other detected Wikipedia links. Therefore, when a term can be mapped to multiple articles, this ambiguity can be resolved by determining the article that best relates to other articles associated with other terms in the input text. This idea was inspired from existing efforts that measure semantic relatedness between Wikipedia links (Ferragina & Scaiella, 2012; Milne & Witten, 2008).The relatedness between any Wikipedia articles can be measured using the following Equation. (Milne & Witten, 2008)

$$\boldsymbol{relatedness}(\mathbf{pa}, \mathbf{pb}) = \mathbf{1} - \frac{\boldsymbol{log}(\boldsymbol{max}(|\boldsymbol{A}|,|\boldsymbol{B}|)) - \boldsymbol{log}(|\boldsymbol{A} \cap \boldsymbol{B}|)}{\boldsymbol{log}(|\boldsymbol{W}|) - \boldsymbol{log}(\boldsymbol{min}(|\boldsymbol{A}|,|\boldsymbol{B}|))} \qquad \textbf{(3.1)}$$

Where $\mathbf{p_a}$ and $\mathbf{p_b}$ are the two articles of Wikipedia, **A** and **B** are the set of all articles that are linked to $\mathbf{p_a}$ and $\mathbf{p_b}$ respectively, and **W** is set of all Wikipedia pages.

Note that the relatedness score gets higher if there are more Wikipedia articles that link to both **pa** and **pb**, i.e. $|\boldsymbol{A} \cap \boldsymbol{B}|$.

For each ambiguous article, we calculate its relatedness score with each candidate article detected in the input text by using Equation 3.2. The overall weight of the ambiguous article is the average of pairwise relatedness scores:

$$\boldsymbol{Weight}_{\mathbf{p}} = \frac{\sum_{i=0}^{n} \boldsymbol{relatedness}(\boldsymbol{p}, \mathbf{Ci})}{|\boldsymbol{C}|} \qquad \textbf{(3.2)}$$

Where **p** is the Wikipedia article for which we need to calculate the average relatedness, **C** is the set of all other candidate articles, $\mathbf{C_i}$ is a candidate article, and **n** is the total number of articles detected from the input text.

Of all Wikipedia articles mapped to a single term, the article with the highest weight is chosen to be linked to the term. This process is repeated for each detected term in the text until selecting the best articles for all terms.

For example, for the input text **"يوروفيجن تحظر رفع علم فلسطين"**, we have two terms **"يوروفيجن"** and **"علم فلسطين".** The term **"علم فلسطين"** related to one Wikipedia article while **"يوروفيجن"** related to **6** artilces. Table 3.3 shows all the articles related to the two

terms and the weight for each article. The article **"يوروفيجن"** was chosen for the term **"يوروفيجن"** because it has the highest weight upon all other articles related to this term.

Table 3.3: Weight of different ambiguous articles

Input Text	Term	Article	Weight
يوروفيجن تحظر رفع علم فلسطين	**يوروفيجن**	مسابقة يوروفيجن	0.0
		مسابقة يوروفيجن للأغاني	0.46
		يوروفيجن	**0.56**
		مسابقه يوروفيجن للأغاني	0.0
		المغرب في يوروفيجن	0.0
		مسابقة يوروفيجن للأطفال	0.0
	علم فلسطين	**علم فلسطين**	**0.51**

- **Terms Filtering**

The article disambiguation phase generates a set of articles to link with candidate terms in the input text, one article per term. These candidate terms; however, have to be filtered to discard the terms that may not be meaningful to the user.

To detect irrelevant terms, we utilized two features that indicate the importance of the associated Wikipedia articles. The first feature is the link probability, which means the probability that the term is used as a link in Wikipedia. The more the term used as a link in Wikipedia articles, the more importance it gains. For any term **a;** the link probability, **P (a),** is calculated as the following:

$$P(a) = \frac{Number\ of\ occurances\ of\ a\ as\ an\ anchor}{Total\ number\ of\ occurances\ of\ a\ in\ Wikipedia} \quad (3.3)$$

Where P (a) is the link's probability

The other feature we used to determine irrelevant terms is the coherence between the term and other terms detected in the short text. Our assumption is that a term gains more importance if it is related to other terms in the short text. In contrast, it will be considered irrelevant or less important if it is not strongly related to the surrounding terms.

The coherence between terms can be determined by measuring the relatedness between their corresponding Wikipedia articles. Therefore, we used the relatedness scores

calculated from the article disambiguation phase. Finally, the filtering score of a term **a** is calculated using the following weighted measure:

$$\mathbf{F(a) = \alpha\, C(a) + \beta\, P(a)} \qquad \mathbf{(3.4)}$$

$$\mathbf{Where\ \alpha + \beta = 1.0}$$

Where **F (a)** is the filtering score that denotes whether the term **a** is significant or not and its value ranges from 0 to 1. **P (a)** indicates the link probability, **C(a)** indicates the coherence score which equals the average relatedness between **a** and all other selected articles of detected terms in the input text.

The factors **α** and **β** decide the contributions of the two scores in the final filtering score. We performed several experiments to determine the best values for **α** and **β**, and found that **α = 0.7** and **β = 0.3** gave the most acceptable results for the underlying dataset (These experiments will be presented in Chapter 4). However, **α** and **β** can have different values depending on the dataset being usedand it is up to the user to set these values. After that, if the filtering score **F (a)** is less than a predefined threshold its annotation will be neglected. Based on our experiments on the collected dataset, threshold was set to **(0.3)** as this value generated the best results.

Note that the relatedness measure in Equation 3.2 was used twice in our annotation approach: the first use chose a Wikipedia article that best represents a term in the input text. This is needed to prune disambiguated articles when a single term matches with multiple Wikipedia articles. The second use of the relatedness measure was filter out terms in the input text that have been linked to a Wikipedia article but may be irrelevant to the user. The computation of the relatedness score in the second time is done faster that the first time as the disambiguate articles are already excluded.
For example, the input text **"دونالد ترامب يفوز بالانتخابات التمهيدية بولاية مشيغان"** have three detected terms **"دونالد ترامب"**, **"مشيغان"**, and **"يفوز"**. Those terms are related to Wikipedia articles. While the terms **"دونالد ترامب"**, and **"مشيغان"** are considered relevant since they have high scores **(0.75)** and **(0.71)** respectively, the term **"يفوز"** considered irrelevant due to its low score **(0.0117)**.

3.4 Case Study

In the following case study, we illustrate a full scenario of the dynamic linking process, showing how the end user query is processed in every step of the approach until the final output is generated. Figure 3.3 shows a simple user interface that enables the user to input the Arabic text and then initiate the annotation process by clicking on the button. Suppose that the user entered the following short text **" الدوري الإيطالي لاتسيو يتغلب على ضيفه انتر ميلان بهدفين نظيفين في الأسبوع 36 من المسابقة"** in the text field.

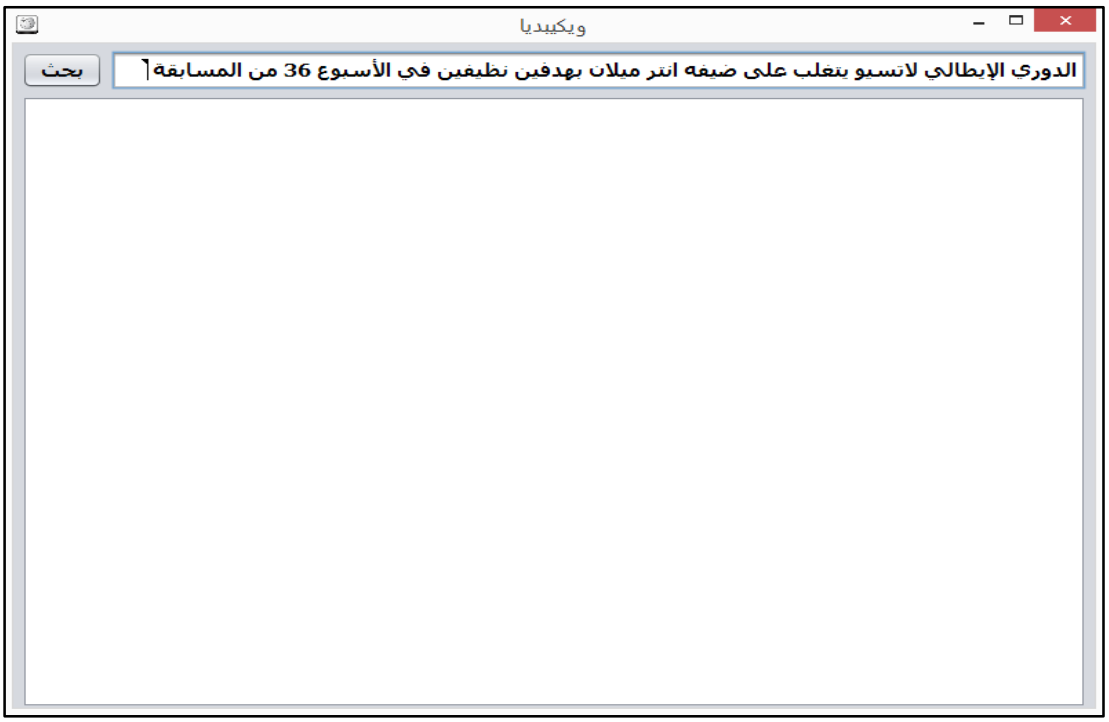

Figure 3.3: Graphical User Interface to enter the short text and show the results

Preprocessing

The input text is first processed by applying stop-word removal, text normalization and tokenization. Afterwards, N-grams are generated from the text where n starts from 1 to 3. Figure 3.4 shows the generated N-grams.

Input Text
الدوري الإيطالي لاتسيو يتغلب على ضيفه انتر ميلان بهدفين نظيفين في الأسبوع 36 من المسابقة

The resulted set of phrases after applying the stop-word removal and word tokenization

N-gram	Resulted Phrase
1-gram	الدوري
	الإيطالي
	لاتسيو
	يتغلب
	انتر
	ميلان
	بهدفين
	نظيفين
	الأسبوع
	المسابقة
2-gram	الدوري الإيطالي
	الإيطالي لاتسيو
	لاتسيو يتغلب
	يتغلب ضيفه
	ضيفه انتر
	انتر ميلان
	ميلان بهدفين
	بهدفين نظيفين
	نظيفين الأسبوع
	الأسبوع المسابقة
3-gram	الدوري الإيطالي لاتسيو
	الإيطالي لاتسيو يتغلب
	لاتسيو يتغلب ضيفه
	يتغلب ضيفه انتر
	ضيفه انتر ميلان
	انتر ميلان بهدفين
	ميلان بهدفين نظيفين
	بهدفين نظيفين الأسبوع
	نظيفين الأسبوع المسابقة

Figure 3.4: Preprocessing of input text to generate the N-gram phrases.

Terms detection

For each n-gram generated from the previous step, we search for all matching Wikipedia articles. The matching process is simply done by searching titles of Wikipedia articles. Note that each n-gram can match with multiple articles. Thus,

each n-gram as well as all Wikipedia articles that match with it are preserved at this stage. N-grams that do not match with any Wikipedia content will be discarded. In our running example, the resulting terms, i.e. n-grams, with all their related Wikipedia articles are shown in Table 3.4. Note that Wikipedia each article in Table 3.4 is identified by a unique ID number.

Table 3.4: Terms Detected for the input text

Phrase	**Related Articles**	
	Article ID	**Article Title**
الدروي الإيطالي	2834588	الدوري الإيطالي الممتاز
	898506	قائمة هدافي الدوري الإيطالي
	22479	الدوري الإيطالي
	16580	الدروي الإيطالي الدرجة الأولى
	1035203	الدوري الإيطالي 2010-11
	2686822	ملحق: هدافو الدوري الإيطالي
	386378	الدوري الإيطالي الدرجة الثانية
	1758362	فضيحة الدوري الإيطالي 2006
	2400246	الدوري الإيطالي 1991-1992
	788203	فضيحه الدوري الإيطالي 2006
	1681192	هدافو الدوري الإيطالي
	2686821	ملحق: قائمة هدافي الدوري الإيطالي
انتر ميلان	264811	نادي انتر ميلان
	88486	انتر ميلان
الأسبوع	408721	نهاية الأسبوع (برنامج)
	885668	نهاية الأسبوع
	2889454	الأسبوع؟
	582350	الأسبوع الذهبي

Phrase	Related Articles	
	Article ID	**Article Title**
	1158471	حظك هذا الأسبوع (فيلم)
	254009	الأسبوع
	1896118	نهاية الأسبوع المفقودة (فيلم)
	819301	أخبار الأسبوع (صحيفة)
	2834885	أخبار الأسبوع (جريدة، الجزائر)
	2830417	الأسبوع المأساوي
	1205299	الأسبوع المأساوي (الأرجنتين)
	1134878	نهاية الأسبوع (عطلة)
	2438621	نهاية الأسبوع (أسبوعية جزائرية)
	1707905	الأسبوع المأساوي (كاتالونيا)
لاتسيو	143752	لاتسيو
	1977314	إحصائيات نادي لاتسيو
	1014117	كولونا، لاتسيو
	39109	نادي لاتسيو

Article Disambiguation

Article disambiguation ensures that each detected term is mapped to the most relevant article among all articles related to this term in Wikipedia. For each term, the selection process is done by calculating the relatedness values between every candidate article associated with that term and the articles associated with all other terms discovered in the input text. Equations **3.1** and **3.2** were used for this purpose. The best Wikipedia article for each term is the one with the highest average relatedness among all other articles related to this term. The following table shows the average relatedness values for each article. Articles with the highest relatedness scores are highlighted in a different color.

Note that some articles have zero relatedness scores.It means that they have no coherence with the other articles denoted by the terms in the input text.

Table 3.5: Weight of the related articles

Phrase	Related Articles	
	Article Title	**Weight**
الدروي الإيطالي	الدوري الإيطالي الممتاز	0.73
	قائمة هدافي الدوري الإيطالي	0.0
	الدوري الإيطالي	0.72
	الدروي الإيطالي الدرجة الأولى	0.55
	الدوري الإيطالي 11-2010	0.0
	ملحق: هدافو الدوري الإيطالي	0.68
	الدوري الإيطالي الدرجة الثانية	0.70
	فضيحة الدوري الإيطالي 2006	0.0
	الدوري الإيطالي 1991-1992	0.0
	فضيحه الدوري الإيطالي 2006	0.0
	هدافو الدوري الإيطالي	0.0
	ملحق: قائمة هدافي الدوري الإيطالي	0.68
انتر ميلان	نادي انتر ميلان	0.63
	انتر ميلان	0.68
لاتسيو	كولونا، لاتسيو	0.0
	لاتسيو	0.62
	إحصائيات نادي لاتسيو	0.71
	نادي لاتسيو	0.73

Terms Filtering

In this step, irrelevant terms are determined by calculating the filtering score explained in term filtering section. Table 3.6 shows the link probability, the coherence score and the overall filtering score for each term from the input text.

Table 3.6: Link probability، coherence score and filtering score of the terms

Article	Link Probability	Coherence Score	Filtering Score
الدوري الإيطالي الممتاز	0.56	0.73	0.69
انتر ميلان	0.64	0.68	0.66
نادي لاتسيو	0.44	0.73	0.64
الأسبوع	0.013	0.0	0.004

Terms that had filtering scores less than the predefined threshold, i.e. **0.3** in our experiment, were discarded. For example, the term **"الأسبوع"** was discarded as it had a filtering score the equals to **0.004**. This result makes sense as this term does not give an important meaning to reader. This results also indicates why the coherence score was weighted higher than the link probability value in the filter. The following figure, Figure 3.5 shows the final wiki terms that occurred to the end user. Each output consists of the term and the link to the Wikipedia article representing it.

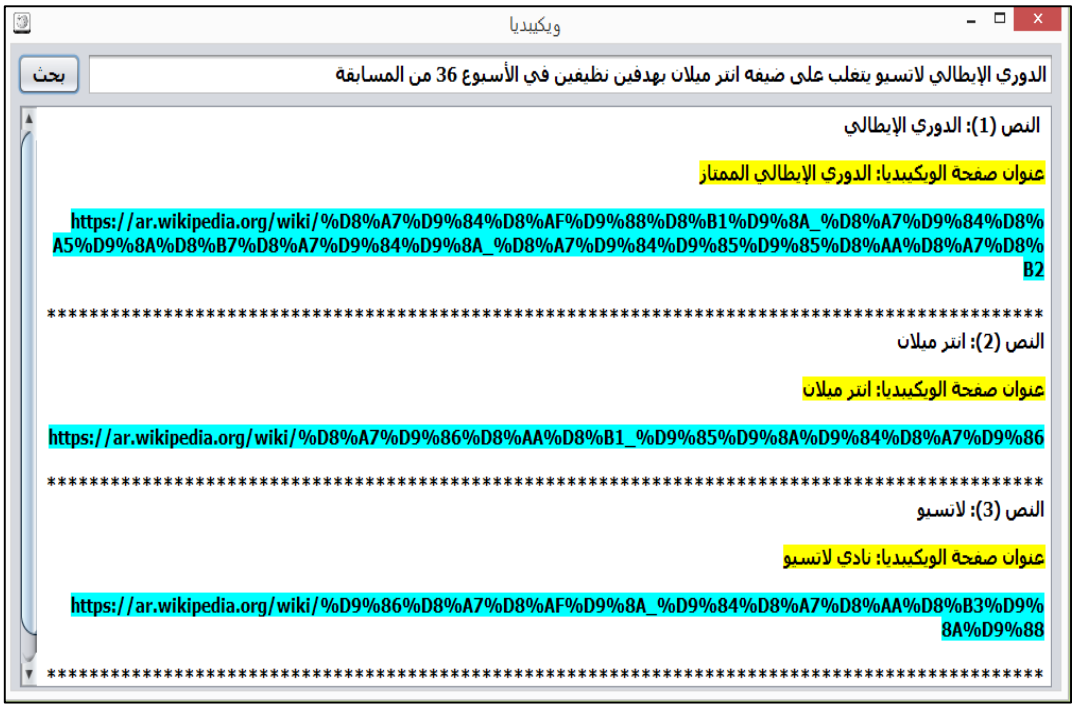

Figure 3.5: The final wiki terms displayed to the user.

3.5 Tools

- **Java Wikipedia Library (*JWPL*)**

JWPL (Java Wikipedia Library) was used to parse and extract information from the XML dump file. ***JWPL*** is an open source Java-based application programming interface that offers free access to all Wikipedia available information. In our research we utilize only JWPL Core (API, and DateMachine) (Zesch et al., 2008). JWPL is available on(Github, 2016)

- **Lucene Search Engine**

Apache Lucene (Lucene, 2016) is an open source java-based search engine library. It offers a high-performance cross-platform full-text solution. We indexed the Wikipedia content by using the Apache Lucene search engine to enable for fast access and search over the Wikipedia content.

- **MySQL**

MySQL (GPL, 2016)is an open source multi-threaded, multi-user Structed Query Language (SQL) database server licensed under the GNU General Public License. MySQL was used to store the content of Wikipedia articles in a local database. Note that all page details are stored in a relational database as an intermediary stage after being extracted from the XML dump file. Important page details such as the title, inlinks and outlinks, which are necessary for the calculations, are indexed by the Lucene search engine and stored in the hard drive.

3.6 Summary

This chapter presents the methodology we followed to construct the desired dynamic linking approach. First XML Wikipedia dump was stored on a local database and the Wikipedia content was indexed to allow for fast access and search over the Wikipedia. Then the annotation process starts with preprocessing of input text by applying stop words removal, tokenization, and n-gram generation. The second step of the annotation is Wikipedia mapper which retireves all Wikipedia articles whose

title contains the detected terms in the input text. Then assures that the detected terms are linked to the most appropriate articles by calculating agreement between the current Wikipedia article and the articles associated with all other terms. The third step of the annotation process is to filter some of the generated terms that may not be meaningful to the user. Two features were used to filter such terms: the first feature is the link probability, which means the probability that the term is used as a link in Wikipedia. And the second feature is the coherence between the term and other terms detected in the short text. Our assumption is that a term gains more importance if it is related to other terms in the short text. In contrast, it considered irrelevant or less important if it is not strongly related to the surrounding terms.

Finally, an example of using the linking approach was discussed in detail. The proposed approach enables the end user to enter the input text, and outputs terms with the best matching Wikipedia articles.

Chapter 4
Results and Discussion

Chapter 4
Results and Discussion

4.1 Introduction

This chapter presents the approach we conducted to evaluate our approach for dynamic linking with the following objectives in mind:

1- Assess the reliability of the dynamic linking approach: we aimed to explore to what extend the proposed approach can accurately link the terms in the input text to the most relevant articles from Wikipedia. We were also interested in exploring any potential errors in linking and the reasons of these errors.
2- Assess the efficiency of the proposed approach: we also aimed to measure the time efficiency of the linking approach and what steps can potentially consume time more than others.

Similar approaches from the state of the art have been often evaluated by being compared with other approaches (Ferragina & Scaiella, 2012; Ren et al., 2013; Shams Eldin & El-Beltagy, 2013). However, we are not aware of any similar approach that utilizes the Arabic version of Wikipedia for dynamic linking to compare with. Therefore, we opted to assess our approach by comparing the output of our approach with the annotations made manually by human subjects over the same dataset.

4.2 Dataset and Human Subjects-opinion

4.2.1 Development Dataset

The dataset was a set of 100 Arabic short texts of different types including tweets, Facebook posts, and Telegram channels' feeds. The short texts were divided as the following: 70 tweets, 21 Telegram feeds, and 9 Facebook posts**.** The source of the gathered data are News, technology, and sports pages and channels: Aljazeera Channel, TRTalarabiya, RTarabic, Sky News Arabia, Aljazeera Sports (AJASports), AJASciTec, AlArabiya, and cnnarabic. Table 4.1 shows a snapshot of the dataset (Refer to appendix A for the complete dataset). The complete dataset can be downloaded from **https://github.com/FatoomMFayad/Dynamic-Linking**

Table 4.1: Snapshot of gathered dataset

No	Text	Source
1	ميسي في مواجهة خيخون رغم الكدمات	Twitter
2	بعد مجلس التعاون الخليجي والجامعة العربية البرلمان العربي يعتبر حزب الله جماعة إرهابية	Twitter
36	مليشيا الحوثي وصالح تقصف أحياء عدة في الجبهة الشرقية من تعز	Telegram
72	توقف نادر لأشهر ساعة في العالم بيج بن تصمت بعد 160 عاماً من الرنين	Facebook
73	تعرض قوات الاحتلال لإلقاء زجاجات حارقة قرب جبل المكبر في القدس المحتلة	Facebook

4.2.2 Evaluation Process

For the assessment of reliability, we run our linking approach over the dataset and recorded the results which consisted of a set of terms, from the input text, and generated links to Wikipedia articles. Results were then given to two human raters to assess them. The two raters were University lecturers and had a long experience in using Arabic Wikipedia. The raters worked independently to undertake two tasks: in the first task, they were asked to inspect the short texts of the dataset to identify terms that can be linked to Wikipedia articles. After identifying these terms, the raters were asked to search Wikipedia for the article that best explains each term.

The results of the rating process was as the following: The first rater linked a total of **225** terms to Wikipedia articles while the second rater linked **215** terms. Both raters agreed over **213**, giving a percentage of agreement which equals the agreed terms divided by the average of the two raters terms **(213/ (225+215)/2) = 96.8%.** The disagreed results included either terms that were linked differently by the two raters **(3 terms)**, or that were linked by one rater but not the other **(7 terms)**. Disagreed results were then discussed and reappraised to reach a mutually acceptable opinion. The final consensus result included a total of **213** terms associated with Wikipedia articles. This final result was used for comparison with output of applying the proposed linking approach over the same dataset.

Table 4.2 depicts how results were collected from the raters: the first column to the left denotes the number of the short text in the dataset. The second column is filled with the term to be linked. The third column is filled with the link to Wikipedia

article.Refer to Appendix B for the full results collected from the raters. Also it can be downloaded from **(https://github.com/FatoomMFayad/Dynamic-Linking)**

Table 4.2: depicts how results were collected from the raters

Text_no	Article Title	URL
6	القاهرة	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%82%D8%A7%D9%87%D8%B1%D8%A9
7	ريال مدريد	https://ar.wikipedia.org/wiki/%D8%B1%D9%8A%D8%A7%D9%84_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF
7	نادي خيتافي	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D8%AE%D9%8A%D8%AA%D8%A7%D9%81%D9%8A
7	نادي برشلونة	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D8%A8%D8%B1%D8%B4%D9%84%D9%88%D9%86%D8%A9
7	أتلتيكو مدريد	https://ar.wikipedia.org/wiki/%D8%A3%D8%AA%D9%84%D8%AA%D9%8A%D9%83%D9%88_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF
8	فرانشيسكو توتي	https://ar.wikipedia.org/wiki/%D9%81%D8%B1%D8%A7%D9%86%D8%B4%D9%8A%D8%B3%D9%83%D9%88_%D8%AA%D9%88%D8%AA%D9%8A
8	نادي روما	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D8%B1%D9%88%D9%85%D8%A7

4.3 Evaluation Metrics

Precision and recall are basic measures used in evaluating how relevant are the retrieved results of an ***IR*** system. This evaluation depends on the comparison of real results and the effective results of the assessed system. Precision and recall are defined as follows(Ting, 2010):

$$\textbf{\textit{Precision}} = \frac{(True\ Positive)}{(True\ Positive + False\ Positive)} \qquad \textbf{(4.1)}$$

$$\textbf{\textit{Recall}} = \frac{(True\ Positive)}{(True\ Positive + False\ Negative)} \qquad \textbf{(4.2)}$$

Based on previous studies (Kulkarni, Singh, Ramakrishnan, & Chakrabarti, 2009)**,** precision and recall were adapted to our experiment as the following:

- True-positive is the number of terms that were linked to relevant Wiki articles.
- False-positive is the number of terms that were linked to non-relevant Wiki articles.

- False-negative is the number of terms that were not linked to any Wiki article despite the presence of relevant articles in Wikipedia.
- True-negative is the number of terms that should not be linked to any Wikipedia article.

Given the above definitions, the precision and real in our experiment are calculated as the following:

$$\textbf{\textit{Precision}} = \frac{\textit{Total number of terms linked to relevant Wikipedia articles}}{\textit{Total number of terms linked to Wikipedia articles}} \quad \textbf{(4.3)}$$

$$\textbf{\textit{Recall}} = \frac{\textit{Total number of terms linked to relevant Wikipedia articles}}{\textit{Total number of terms that can be linked to Wikipedia articles}} \quad \textbf{(4.4)}$$

We calculated the precision and recall for each short text in the input text. We also assessed the approach by using the F-measure, the harmonic mean of precision and recall. This mean is approximately the average of the two when their values are close(Sasaki, 2007).

$$\textbf{\textit{F-measure}} = \frac{2*\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad \textbf{(4.5)}$$

We also calculated accuracy and error rate of the approach using the following equations(Zhu, Zeng, & Wang, 2010):

$$\textbf{\textit{Accuracy}} = \frac{(\textit{True Negative} + \textit{True Positive})}{(\textit{True Positive} + \textit{False Positive} + \textit{True Negative} + \textit{False Negative})} \quad \textbf{(4.6)}$$

$$\textbf{\textit{Error Rate}} = \frac{(\textit{False Negative} + \textit{False Positive})}{(\textit{True Positive} + \textit{False Positive} + \textit{True Negative} + \textit{False Negative})} \quad \textbf{(4.7)}$$

Table 4.3 shows precision, recall and F-measure for a sample of the input texts. Refer to Appendix C for the full results of all short texts of the dataset and how we calculated the precision and recall.

Table 4.3: Precision, Recall, and F-measure

Text	Precision	Recall	F-measure
ميسي في مواجهة خيخون رغم الكدمات	1	1	1
بعد مجلس التعاون الخليجي والجامعة العربية البرلمان العربي يعتبر حزب الله جماعة إرهابية	0.4	0.5	0.44
الجزيرة الإماراتي يسعى لضم توتي	1	1	1
أوباما: نسعى إلى تحسين وضع اللاجئين وتقديم حياة أفضل لهم	0.67	1	0.8

4.4 Results and Discussion

Table 4.4**:** Evaluation metrics of the system

Number of Resulted Terms	**235**
Macro Average Precision	**74.70 %**
Macro Average Recall	**82.63 %**
F-measure	**75.21%**
Accuracy	**71.79 %**
Error Rate	**28.21 %**

Results are presented in Table 4.4 The proposed approach generated **235 terms**. It achieved **(71.79%)** accuracy, **(74.70%)** macro average precision, and **(82.63 %)** macro average recall; while the error rate was **(28.21%)**. To further explain our results, we inspect the results thoroughly to identify the main sources of errors and differences between the system output and the base links. For simplicity, we refer to the links given by human subjects as the base links. Errors can be classified into the following categories based on the source of errors:

- **Errors due to Article disambiguation:** A considerable number of mismatches between the base links and the generated links were attributed to the variability of the disambiguation process. In some cases, article disambiguation resulted in articles that were more specific and focused than the articles pointed by the base links. For example, in the text **" ضغوط أمريكية على آبل لفتح آيفون في الجرائم"** the word **"أبل"** was associated with the article titled as **"مطورو أبل"** rather than the article titled as **"أبل"**. In other examples, the word **"الكرملين"** was linked to article **"كأس الكرملين"**, the word **"الفلوجة"** was linked to an article with the title: "أحداث الفلوجة", and the word **"أمستردام"** was mapped to an article titled as: **"مؤشر بورصة أمستردام"**. In all previous examples, the disambiguation process picked articles that were more specific rather than choosing more general ones. From the system's point of

view, some of these results can be explained by the fact that specific articles could have higher relatedness scores than the general articles. This is because specific articles often extend the content of more general articles, and thus they are still coherent with the rest of terms in the input text. This was the most common type of errors, contributing with **60.6 %** of the total number of errors generated. Surprisingly, the tendency to choose specific articles over general ones, sometimes resulted in more accurate results compared to base links. For example, the word **"شيعة"** in the text: **"14 قتيلاً في تفجير استهدف زوار شيعة بالعراق"** was mapped to the Wikipedia article titled as **"شيعة العراق"**. This result is more convenient that the base link to the article titled as **"الشيعة"**. Similarly, the phrase **"اللاجئون الفلسطينيون"** in the text **" اللاجئون الفلسطينيون في مخيم اليرموك بين مطرقة حصار النظام وسندان المواجهات بين تنظيم الدولة و جبهة النصرة"** was associated with the article **" اللاجئون الفلسطينيون في سوريا"** instead of the article **"اللاجئون الفلسطينيون"** as indicated in the base links. In another example, the term **"تنظيم الدولة"** in the text:**" أميركا ترسل قوات إضافية لمواجهة تنظيم "الدولة في العراق""** was associated to the article titled as **"القاعدة في العراق"**. While this result is different from the standard result which is the article titled as **"تنظيم الدولة"**. It looks more related to the context of the text. These results are more related to the context of the text than the base links.

One solution to the problem of selecting specific articles is to discard articles with titles that are longer than the target terms. For example, the article titled as **" مؤشر بورصة أمستردام"** should be discarded as its title is three-word length while the target term is a single word. This solution; however, can cause other errors such as excluding articles with the titles: **"جمهورية مصر العربية"** and **"الولايات المتحدة الأمريكية"** for the words**: "مصر"** and **"أمريكا"** respectively. A potentially more convenient, though computationally expensive approach will be to consider the content of articles to resolve ambiguity.

- **Errors due to term filtering:** As explained in Section 3.3.2, the link filtering aims to discard generated links that do not convey important meanings to the reader. However, two problems with the filtering approach resulted in the pruning of detected links that should be preserved. The first problem is the low computed filtering scores, thereby causing some important terms to be discarded. For

example, the word "**أردوغان**" in the text "**أردوغان: المنظمات الإرهابية التي تدعي الإسلام أضرت بالإسلام أكثر من الأعداء**" was discarded although it was part of the base links. This result can be explained by the low coherence between the word "**أردوغان**" and the other terms in the text. Therefore, the word was excluded since its filtering score fell under the predefined threshold. In contrast, some terms that were considered irrelevant by our human subjects were detected and associated with Wikipedia articles by our approach. These terms had filtering scores slightly over the threshold of **0.3**. Considering the following short text as an example "**مقال بمجلة نيوزويك الأميركية يتساءل عما فعله الرئيس باراك أوباما تجاه جرائم نظام الرئيس السوري بشار الأسد**": the word "**مقال**" was linked with a Wikipedia article titled as "**مقال**", and the word "**الرئيس**" was linked with another article titled as "**الرئيس الأمريكي**". While these generated links seem consistent with the context of the text, they were dismissed by the human subjects as they were considered less important. These results triggers the need to perform more experiments to determine the best values for the threshold and the weights of the link probability and coherence when calculating the filtering score. These errors; however, do not underestimate the value added by the filtering step which made the results much reasonable and less overwhelming for end users. To give a realistic example on the positive impact of the filtering step, consider the following text "**قوقل تخطط لعرض الموضوعات الأكثر تداولًا في مربع البحث**": Our linking approach detected only a single term in this text, which is "**قوقل**" and linked it to the relevant article. This result matched the links detected by the human subjects. After disabling the filter, the words "**قوقل, تخطط, موضوعات, تداول, بحث**" were all linked to articles from Wikipedia.

- **Errors due to lack of semantic reasoning:** The proposed linking approach primarily relies on the syntactic matching with the titles of Wikipedia articles as well as the link structure to determine relatedness between articles. However, the lack of semantic inference may result in results that mismatch with the user's interests. For example, in the text"**لوبان يتنبأ لابنته بالفشل في انتخابات الرئاسة**", the word "**لوبان**" was linked to the article of Le Pen the daughter, i.e. Marine Le Pen, while it should be linked to the article of Le Pen the father, i.e. Jean Marie Le Pen, according to the context of the text. In another example" **باعتباره الخليفة المحتمل للرئيس بوتفليقة ... هل يقوم الوزير السابق شكيب خليل بحملة انتخابية مبكرة**"، the word "**الوزير**" was

associated with the article of "خليل الوزير" although this result conflicts with the intended meaning of the word.

4.5 Time Efficiency

We evaluated the execution time of the 100 short texts of the dataset. The specifications of the machine used in the evaluation process is shown in Table 4.5.

Table 4.5: Evaluation Machine Specification

Processor Type	**Intel® core i5-5200u**
Processor Clock Speed	**2.20 Giga Hertz**
Installed Memory	**8 Giga Byte**
Operating System	**Windows 8.1 Enterprise**
System Type	**64 bit operating system**

Figure 4.1 depicts the execution times for the 100 short texts. Table 4.6 summarizes the results. The average execution time for the 100 short texts was **40.16 seconds** and the standard deviation was **33.01**. The minimum execution time for any text was **5.927 seconds** and the maximum execution time was **235.627 seconds**. It is obvious from Figure 4.1 that execution times varied largely (**SD = 33.01**).

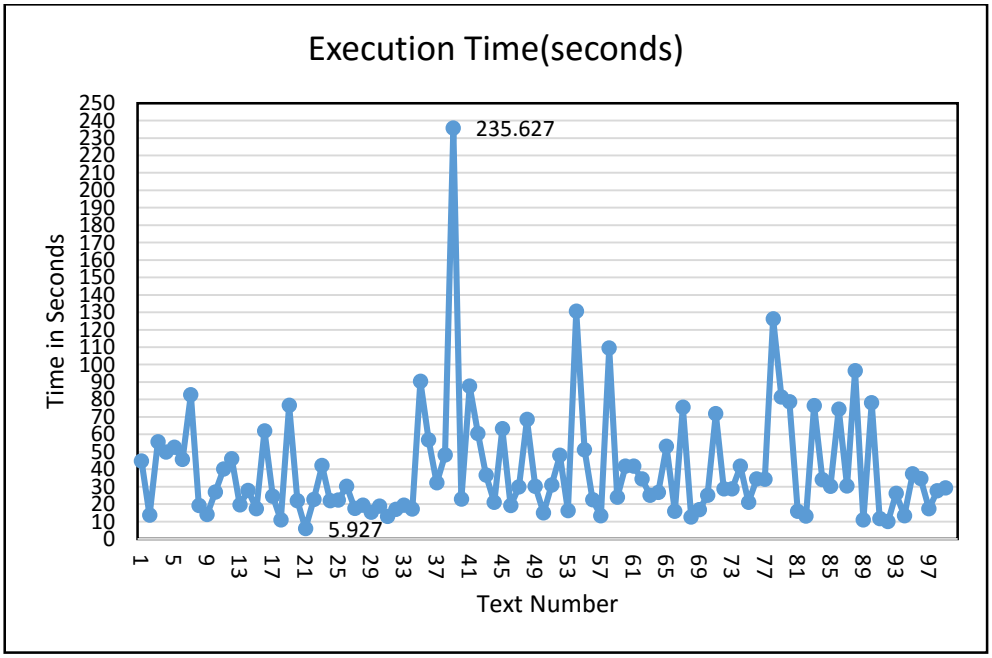

Figure 4.1: Execution Time for the 100 short texts

Table4.6: Execution Time

Average Execution Time	**40.16 seconds**

Minimum Execution Time	**5.927 seconds**
Maximum Execution Time	**235.627 seconds**
Standard Deviation	**33.01**

To further explain these results, we measured the time required to execute each of the following steps: the mapping, the disambiguation process and the filtering. Table 4.7 shows the time required for each of the steps.

Table 4.7: Average Execution Time of the Detailed Steps

Step	Average Execution Time	Standard Deviation
Mapping	**21.84 seconds**	**11.25**
Disambiguation	**23.32 seconds**	**39.66**
Filtering	**1.56 seconds**	**1.6**

These results indicate that the disambiguation step consumed the longest time, followed by the mapping step and filtering step. The time for the mapping step had a low variance across the 100 texts, meaning that it does not significantly vary across different texts. However, the long time required for the mapping process was due to the number of search queries on Wikipedia which was equal to the number of n-grams generated from the input text.

The disambiguation step consumes the longest time as compared with other steps. In addition, the time required to disambiguate pages varied largely across the short texts. For example, one text took only **1.44** seconds to complete the disambiguation phase, while another step took about **87.08** seconds to complete the same phase. This high variance can be explained by the number of ambiguous articles obtained from the mapping step. The larger the number of ambiguous articles, the longer time is needed to handle them. The long time for the disambiguation phase was due to the fact that the disambiguation measure is a pairwise measure whereas the relatedness is calculated between every pair of articles. This means that the complexity of the disambiguation measure increases factorially as the number of Wikipedia matches increases.Recall that Equation 3.1 entails calculating the pairwise relatedness between candidate articles. Considering that some terms was mapped to several tens of Wikipedia articles, a large number of calculations should be done. For example, the term "**Saudi Arabia**" **"السعودية"** in the short text **"سناب شات لمراقبة السائقين المخالفين في السعودية"** was mapped to **405** articles. For each of these articles, the pairwise

relatedness with every article linked to other terms should be calculated. In contrast, some texts required shorter times to resolve disambiguation due to the small number of ambiguous articles obtained from the mapping step. The analysis of the disambiguation step explains the high fluctuation and variance in the overall execution times of the 100 short texts.

The filtering step consumed the shortest time in comparison with the other steps. Although the filtering step is similar to the disambiguation step in terms of relying on the relatedness measure to estimate the coherence, this step is much faster than the disambiguation step because all ambiguous articles are filtered out, and thus are not considered in the filtering step.

Note that in our experiments we did not use any mechanism for parallelism or execution over a cluster of computing nodes as done in many of previous efforts. Experiments were conducted on a standalone machine.

4.6 Evaluating the Parameters of the Filtering Score

As discussed in Section3.3.2, the filtering process uses a measurement that balances between two features: the link probability and the coherence. In our experiment, the coherence score was weighted more than the link probability (**α = 0.7**, **β= 0.3**). The values of the parameters **α** and **β** were optimized to achieve the best performance. However, these parameters may need to be adjusted based on the nature of the dataset being used. In this section we explore how the performance, in terms of F-measureof the approach, is affected by adjusting the values of **α** and **β**. Since the values of **α** and **β** should sum to 1, the linking approach was tested over the same dataset while setting **α** to the values from 0 to 1 with a step of 0.2. Four iterations were carried out with the settings shown in Table 4.7.

Table 4.8: Iterations of α

Iteration	**α**	**β**
1	0.1	0.9
2	0.3	0.7
3	0.7	0.3
4	0.9	0.1

Figure 4.2 shows how the performance was changed when changing **α** (α = 0.5 was ignored because it returns results that are very close to **α** = 0.7). The results show that the best performance was achieved when **α** was about 0.7**. This value indicates that** the coherence was weighted higher than the link probability in our experiment.

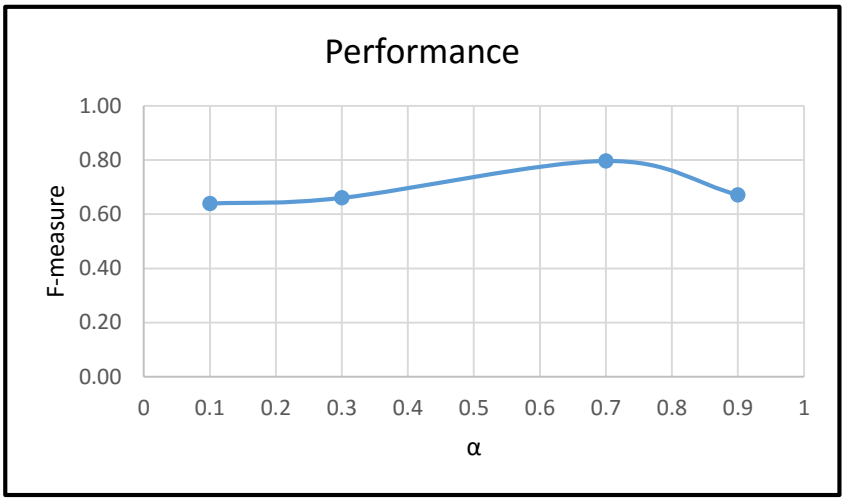

Figure 4.2: Effect of α on Performance

Finally we analysed the effect of varying threshold value on the performance of the approach. The F-measure score was computed on the same dataset while varying the threshold value from 0 to 0.8 with a step of 0.2. Table 4.8 and Figure 4.3 show the impact of different values of threshold on F-measure. Results showed that the best performance was achieved when the threshold was between 0.2 and 0.4. Then, the performance started degrading as the threshold increased. Therefore, the value of 0.3 was chosen for our threshold, meaning that all articles with filtering scores below this value will be discarded.

Table 4.9: Effect of varying threshold values on performance

Threshold	Precision	Recall	F-measure
0	0.46	0.92	0.60
0.2	0.67	0.97	0.76
0.4	0.79	0.93	0.82
0.6	0.53	0.39	0.41
0.8	0.37	0.15	0.21

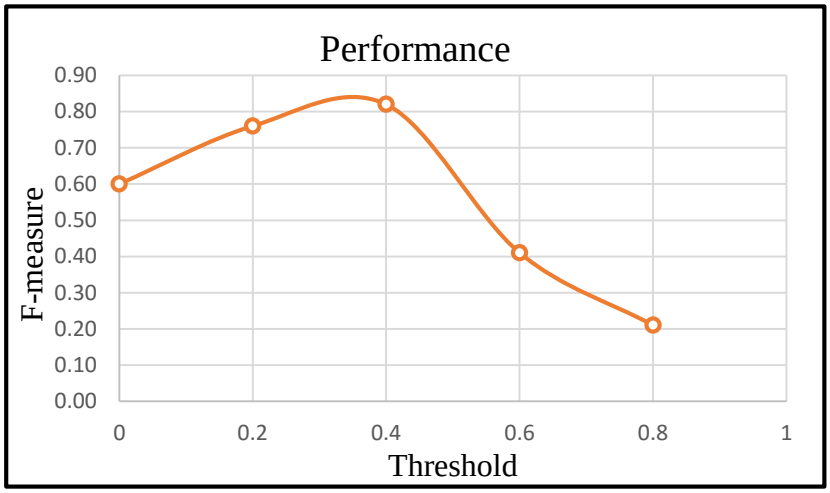

Figure 4.3: Effect of threshold on Performance.

4.7 Summary

This chapter presents the evaluation of the approach. It also discusses the results and the sources of resulting errors (the system assessed using the relevant measures of ***IR)***.

We claim that there is no previous efforts existing in dynamic linking between short texts and Wikipedia in Arabic domain. We formulated a dataset of 100 short texts to assess the approach then the results compared to human subjects' opinion results. The results indicated that our system achieve a high relevant measures with **71.79** accuracy, **74.70** precision, and **82.63** recall.

Chapter 5
Conclusions

Chapter 5
Conclusions

In this thesis we developed a dynamic linking approach for unstructured Arabic short texts by exploiting Arabic Wikipedia as a knowledge source. Given an input short text the approach searches the Wikipedia content for the articles that best describes the most significant terms within the text. The linking approach consists of three subsequent steps: 1) matching the input text with Wikipedia articles. This step aims to identify the set of all possible Wikipedia articles for each term in the input text. 2) Article disambiguation: Given the potentially-large number of candidate articles obtained from the matching step. This step assures that each detected term will be linked to the most relevant article among the candidate articles 3) Article Filtering, which aims to filter out some of the generated terms that may not be meaningful to the user.

The dynamic linking approach was evaluated over a dataset of 100 short texts gathered online from many online Arabic articles. The results of the approach are compared with the results generated by two human subjects. The system was assessed based on the evaluation metrics precision, recall, error rate and accuracy. The results indicated that our system achieved a high relevant measures with **71.79%** accuracy, **74.70** precision, and **82.63** recall. The time efficiency of the system was assessed using a local machine and the average execution time was **39.91** seconds.

The work in this thesis has the following research contributions:

To our knowledge, this is the first work that explored the annotation of Arabic text with links to Wikipedia content. Arabic Wikipedia has only been exploited recently by the Arab computer researchers and few efforts from the literature have tried to interface to the Arabic version of Wikipedia for different purposes such as determining relations between topics (Kanan et al., 2015), named entity recognition (Althobaiti et al., 2014)and ontology generation (Al-Rajebah & Al-Khalifa, 2014).

Our work builds on the techniques used in counterparts efforts that tackled the English and Latin versions of Wikipedia, and tried to adapt them to the Arabic version

of Wikipedia. While existing efforts on Wikipedia-based annotation have largely benefited from the advancements and off-the-shelf tools in support of the processing English Wikipedia, the processing of Arabic Wikipedia for search and annotation in our approach was carried out from scratch.

We proposed an in-depth evaluation of our linking approach and explored the potential shortcomings and strengths of each involved process. This detailed evaluation can inform Arab researchers with the various design options and recommendations for designing similar approaches. The evaluation has also highlighted several issues that have not been addressed in existing efforts such as issues related to article disambiguation and filtering.

The work on this thesis is only a first step towards utilizing Arabic Wikipedia for dynamic annotation and there are many dimensions to extend it:

First, we will consider the time efficiency of the approach. Possible solutions to speed up the linking approach include: a) Explore the use of better matching algorithms, and try to filter generated n-grams before matching them with Wikipedia content. b) Exploit parallel processing and multi-processor techniques. c) Explore faster approaches to resolve ambiguity among candidate articles.

Second, we are trying to improve the article disambiguation approach to produce more accurate and intelligent results. Potential investigations may include: a) Exploit the content of documents when measuring the relatedness between documents (e.g. info boxes, sub-titles) besides the Wikipedia link structure. b) Exploit background knowledge such as WordNet and ontologies to boost the documents' relatedness measure. C) Exploit natural language processing techniques such as named entity recognition as this will help identify the potential categories of words to be annotated.

Third, we are trying to deploy our linking service and make it available for public use. We are currently in the process of building a web service that will enable user to post a short text as input and retrieve annotations as output. Opening the service for the large-scale Arab community will provide valuable info about the potential affordances, shortcoming and future improvements.

References

Abdeen, M., & Tolba, M. F. (2010). *Challenges and design issues of an Arabic web crawler.* Paper presented at the International Conference on Computer Engineering and Systems (ICCES), Cairo,Egypt.

Adnan, M., Warren, J., & Orr, M. (2013). *SemLink—Dynamic generation of hyperlinks to enhance patient readability of discharge summaries.* Paper

presented at the 26th International Symposium on Computer-Based Medical Systems (CBMS), Porto, Portugal.

Al-Rajebah, N. I., Al-Khalifa, H., & Al-Salman, A. (2011). *Exploiting Arabic Wikipedia for automatic ontology generation: a proposed approach.* Paper presented at the International Conference on Semantic Technology and Information Retrieval (STAIR), Kuala Lumpur, Malaysia.

Al-Rajebah, N. I., & Al-Khalifa, H. S. (2014). Extracting ontologies from Arabic Wikipedia: A linguistic approach. *Arabian journal for Science and Engineering, 39*(4), 2749-2771.

Al-Shalabi, R., Kanaan, G., Jaam, J. M., Hasnah, A., & Hilat, E. (2004). *Stop-word removal algorithm for Arabic language.* Paper presented at the Proceedings of 1st International Conference on Information & Communication Technologies: from Theory to Applications, CTTA'04, Damascus, Syria.

Alotaibi, F., & Lee, M. (2013). *Automatically developing a fine-grained arabic named entity corpus and gazetteer by utilizing wikipedia.* Paper presented at the Proceedings of the Sixth International Joint Conference on Natural Language Processing, Nagoya, Japan

Althobaiti, M., Kruschwitz, U., & Poesio, M. (2014). *Automatic Creation of Arabic Named Entity Annotated Corpus Using Wikipedia.* Paper presented at the In Proceedings of the Student Research Workshop at the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL), Gothenburg, Sweden.

Badie, R., Azimzadeh, M., Zahedi, M. S., & Samuri, S. (2014). *Automatic evaluation of search engines: Using webpages' content, web graph link structure and websites' popularity.* Paper presented at the Seventh International Symposium on Telecommunications, Tehran, Iran.

Bahanshal, A. O., & Al-Khalifa, H. S. (2013). *Toward Recipes for Arabic DBpedia.* Paper presented at the Proceedings of International Conference on Information Integration and Web-based Applications & Services, Vienna, Austria.

Barrena, A., Donostia, B. C., Soroa, A., & Agirre, E. (2015). Combining Mention Context and Hyperlinks from Wikipedia for Named Entity Disambiguation. *Lexical and Computational Semantics (* SEM 2015)*, 101.

Chan, E. C., Baciu, G., & Mak, S. (2009). *Cognitive location-aware information retrieval by agent-based semantic matching.* Paper presented at the 8th IEEE International Conference on Cognitive Informatics, Hong Kong.

Cosh, K. (2013). *Text mining wikipedia to discover alternative destinations.* Paper presented at the 10th International Joint Conference on Computer Science and Software Engineering (JCSSE), Khon Kaen, Thailand.

Demartini, G., Firan, C. S., Iofciu, T., Krestel, R., & Nejdl, W. (2008). *A model for ranking entities and its application to wikipedia.* Paper presented at the Web Conference, 2008. LA-WEB'08, Latin American.

Derczynski, L., Ritter, A., Clark, S., & Bontcheva, K. (2013). *Twitter Part-of-Speech Tagging for All: Overcoming Sparse and Noisy Data.* Paper presented at the RANLP, Hissar, Bulgaria.

Farghaly, A., & Shaalan, K. (2009). Arabic natural language processing: Challenges and solutions. *ACM Transactions on Asian Language Information Processing (TALIP), 8*(4), 14-33.

Ferragina, P., & Scaiella, U. (2010). Fast and accurate annotation of short texts with Wikipedia pages. *arXiv preprint arXiv:1006.3498*.

Ferragina, P., & Scaiella, U. (2012). Fast and Accurate Annotation of Short Texts with Wikipedia Pages. *Software, IEEE, 29*(1), 70-75.

Geiß, J., Spitz, A., & Gertz, M. (2015). *Beyond Friendships and Followers: The Wikipedia Social Network.* Paper presented at the Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015, Paris, France.

Github. (2016). JWPL. 1.1.0. Retrieved February 18th, 2016, from https://github.com/dkpro/dkpro-jwpl

GPL. (2016). MySQL. 5.7.9. Retrieved June18th ,2016, from http://dev.mysql.com/downloads/mysql/

Gruetze, T., Kasneci, G., Zuo, Z., & Naumann, F. (2014). *Bootstrapping Wikipedia to answer ambiguous person name queries.* Paper presented at the 30th International Conference on Data Engineering Workshops (ICDEW), Chicago, USA.

Ikikat, F. Y., Gurhan, B., & Diri, B. (2015). *Automatic linking of wikipedia pages by their semantic similarity.* Paper presented at the International Symposium on Innovations in Intelligent SysTems and Applications (INISTA), Madrid, Spain.

Kanan, T., Ayoub, S., Saif, E., Kanaan, G., Chandrasekarar, P., & Fox, E. (2015). Extracting Named Entities Using Named Entity Recognizer and Generating Topics Using Latent Dirichlet Allocation Algorithm for Arabic News Articles: Department of Computer Science, Virginia Polytechnic Institute & State University.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons, 53*(1), 59-68.

Kulkarni, S., Singh, A., Ramakrishnan, G., & Chakrabarti, S. (2009). *Collective annotation of Wikipedia entities in web text.* Paper presented at the Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, Paris, France.

Lucene, A. (2016). Apache Lucene. Retrieved May 17th, 2016, from http://archive.apache.org/dist/lucene/java/

Milne, D., & Witten, I. H. (2008). *Learning to link with wikipedia.* Paper presented at the Proceedings of the 17th ACM conference on Information and knowledge management, California, USA.

Mirizzi, R., Ragone, A., Di Noia, T., & Di Sciascio, E. (2010). Semantic tag cloud generation via DBpedia. In F. Buccafurri & G. Semeraro (Eds.), *E-Commerce and Web Technologies* (pp. 36-48). Boston, MA: Springer US.

Moro, A., Raganato, A., & Navigli, R. (2014). Entity linking meets word sense disambiguation: a unified approach. *Transactions of the Association for Computational Linguistics, 2*, 231-244.

Nakamura, T., Shirakawa, M., Hara, T., & Nishio, S. (2014). *Semantic Similarity Measurements for Multi-lingual Short Texts Using Wikipedia.* Paper presented at the The 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), Warsaw, Poland.

Navigli, R., & Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence, 193*, 217-250.

Passant, A. (2010). *Measuring Semantic Distance on Linking Data and Using it for Resources Recommendations.* Paper presented at the AAAI spring symposium: linked data meets artificial intelligence, California, USA.

Ramudu, B., & Murty, M. N. (2012). *Topic based semantic clustering using Wikipedia knowledge.* Paper presented at the 2012 International Conference on Data Science & Engineering (ICDSE), Kerala, India.

Ren, Z., Van Dijk, D., Graus, D., Van Der Knaap, N., Henseler, H., & De Rijke, M. (2013). *Semantic linking and contextualization for social forensic text analysis.* Paper presented at the Intelligence and Security Informatics Conference (EISIC), 2013 European, Uppsala, Sweden.

Sasaki, Y. (2007). The truth of the F-measure. *Teach Tutor mater*, 1-5.

Scheau, C., Rebedea, T., Chiru, C., & Trausan-Matu, S. (2010). *Improving the relevance of search engine results by using semantic information from wikipedia.* Paper presented at the Roedunet International Conference (RoEduNet), 2010 9th, Sibiu, Romania.

Shams Eldin, S., & El-Beltagy, S. R. (2013). *News auto-tagging using Wikipedia.* Paper presented at the 2013 9th International Conference on Innovations in Information Technology (IIT), Al Ain, UAE.

Sriharee, G. (2014). An ontology-based approach to auto-tagging articles. *Vietnam Journal of Computer Science, 2*(2), 85-94.

Technology, H. K. U. o. S. a. (2014). Web Technologies and Applications. In L. Chen, Y. Jia, T. Sellis & G. Liu (Eds.), *Web Technologies and Applications: 16th Asia-Pacific Web Conference, APWeb 2014, Changsha, China, September 5-7, 2014. Proceedings* (pp. 5-7). Boston, MA: Springer.

Ting, K. M. (2010). Precision and Recall. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning* (pp. 781-781). Boston, MA: Springer US.

Vivaldi, J., & Rodriguez, H. (2014). *Arabic medical terms compilation from Wikipedia.* Paper presented at the 2014 Third IEEE International Colloquium in Information Science and Technology (CIST), Tetuan, Morocco.

Wikimedia. (2015). Wikimedia. Retrieved October 11th ,2015, from https://stats.wikimedia.org/EN/TablesWikipediaAR.htm

Zaraket, F., & Jaber, A. (2013). *MATAr: Morphology-based Tagger for Arabic.* Paper presented at the 2013 ACS International Conference on Computer Systems and Applications (AICCSA), Fes/Ifrane, Morocco.

Zesch, T., Müller, C., & Gurevych, I. (2008). *Extracting Lexical Semantic Knowledge from Wikipedia and Wiktionary.* Paper presented at the LREC, Marrakech, Morocco.

Zhang, H., Zhang, S., Wu, Z., Huang, L., & Ma, Y. (2014). *Predicting Wikipedia Editor's Editing Interest Based on Factor Graph Model.* Paper presented at the 2014 IEEE International Congress on Big Data (BigData Congress), Alaska, USA.

Zhiqiang, L., Werimin, S., & Zhenhua, Y. (2009). *Measuring semantic similarity between words using wikipedia.* Paper presented at the International Conference on Web Information Systems and Mining, 2009. WISM 2009., Shanghai, China.

Zhu, W., Zeng, N., & Wang, N. (2010). Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS® implementations. *NESUG proceedings: health care and life sciences, Baltimore, Maryland*, 1-9.

Appendices

Appendices

Appendix A: Arabic Short Texts Dataset

No	Text	Source
1	ميسي في مواجهة خيخون رغم الكدمات	Twitter
2	بعد مجلس التعاون الخليجي و الجامعة العربية البرلمان العربي يعتبر حزب الله جماعة إرهابية	Twitter
3	الجزيرة الإماراتي يسعى لضم توتي	Twitter
4	أوباما : نسعى إلى تحسين وضع اللاجئين وتقديم حياة أفضل لهم	Twitter
5	اللاجئون الفلسطينيون في مخيم اليرموك بين مطرقة حصار النظام وسندان المواجهات بين تنظيم الدولة و جبهة النصرة	Twitter
6	فايننشال تايمز: آلاف المحتجين عادوا إلى شوارع القاهرة في تحد واضح للوجود الأمني الكثيف، تماما كما حدث في عام 2011	Twitter
7	الليغا : ريال مدريد يسحق خيتافي ويعزز ضغوطه على برشلونة و أتليتكو	Twitter
8	توتي ينقذ روما من الخسارة في أرض مضيفه أتالانتا ويمكنه من التعادل معه بثلاثة أهداف لكل منهما	Twitter
9	أتلتيكو مدريد يفوز و يشدد الخناق على برشلونة	Twitter
10	مانشستر يونايتد الأكثر إنفاقا على وكلاء اللاعبين	Twitter
11	غرناطة يقسو على الضيف ليفانتي بخماسية كاملة في الدوري الإسباني	Twitter
12	زيدان : رونالدو بحاجة إلى الراحة والغياب عن بعض المباريات إذا رغب في إكمال الموسم الشاق للريال	Twitter
13	ملتقى مغردون في الرياض يجمع خمسة وزراء يستخدمون تويتر بالجمهور	Twitter
14	أندرويد إن قادم مع مزيدٍ من التطبيقات لدعم الواقع الافتراضي	Twitter
15	ضغوط أمريكية على آبل لفتح آيفون في الجرائم	Twitter
16	قوقل تخطط لعرض الموضوعات الأكثر تداولًا في مربع البحث	Twitter
17	أردوغان : المنظمات الإرهابية التي تدعي الإسلام أضرت بالإسلام أكثر من الأعداء	Twitter
18	الجيش العراقي يقصف الفلوجة ويقتل مسلحين في نينوى	Twitter
19	هل نهاية نظارات قوقل باتت قريبة	Twitter
20	زعيم المعارضة في الكيان الإسرائيلي هرتسوغ تعليقا على نفق كتائب القسام : محمد الضيف لا يحسب ل نتنياهو أي حساب	Twitter
21	بوتن و أوباما يبحثان سوريا.. الهدنة بخطر والمعارضة تضغط	Twitter
22	جوكوفيتش يتفوق على ميسي	Twitter
23	توتنهام يقلص الفارق مع ليستر بفوزه على ستوك	Twitter
24	أميركا ترسل قوات إضافية لمواجهة تنظيم الدولة في العراق	Twitter
25	اليونسكو تعتمد مصطلح المسجد الأقصى وترفض جبل الهيكل	Twitter
26	إسبانيا تعتقل مغربيا للاشتباه في صلته ب داعش	Twitter
27	مقتدى الصدر يهدد بأنه سيأمر أتباعه بالخروج بمظاهرة مليونية الاثنين المقبل	Twitter
28	ناسا تحتفل بمرور 26 سنة على إطلاق تلسكوب هابل	Twitter
29	الخلافات تطغى على اليوم الثاني من مباحثات اليمن	Twitter
30	أوباما يطلب لقاء خامنئي في رسالتين سريتين	Twitter
31	بعد واتساب .. فايبر تشفر محادثات مستخدميها	Twitter
32	الحوثيون يطلبون وقف المشاورات لمناقشة الهدنة	Twitter
33	قتلى بتفجير سيارتين ملغومتين في بغداد	Twitter
34	ثنائية يوفيتيتش تقود الإنتر للفوز 3-1 على أودينيزي	Twitter
35	إسرائيل تثبّت حكما بسجن الشيخ رائد صلاح	Twitter

No	Text	Source
36	مليشيا الحوثي وصالح تقصف أحياء عدة في الجبهة الشرقية من تعز	telegram
37	الحسم يتأجل والتعادل السلبي يفرض نفسه بين مانشستر سيتي وضيفه ريال مدريد في ذهاب نصف نهائي دوري أبطال أوروبا	telegram
38	قتيلان وعشرات الجرحى جراء غارات على قرى جبل الزاوية بريف إدلب	telegram
39	مقتل 6 مدنيين في قصف للجيش العراقي على الفلوجة ومحيطها	telegram
40	بعد أن أظهر قدرة فائقة على استخدام وسائل التواصل الاجتماعي .. الولايات المتحدة تفتح جبهة جديدة للحرب ضد تنظيم الدولة	Twitter
41	البرلمان العراقي يوافق على استقالة وزير الخارجية إبراهيم الجعفري	telegram
42	التعادل سيد الموقف في مبارة ريال مدريد و مانشستر سيتي والحسم يتأجل	telegram
43	العارضة تحرم ريال مدريد من إحراز أول أهدافه برأسية مهاجمه خيسي	telegram
44	لاجئ سوري يحمل شعلة الأولمبياد في اليونان	telegram
45	اليونيسكو : تدمر احتفظت بأغلب آثارها رغم الأضرار	Twitter
46	أتلتيكو مدريد يهزم ضيفه بايرن ميونيخ بهدف دون رد أحرزه نيغيز في الدقيقة 11 بذهاب نصف نهائي دوري أبطال أوروبا	telegram
47	فوز المدهون بالبوكر .. أسئلة النكبة و الهولوكوست	Twitter
48	المتصدر يوفنتوس يحقق فوزا صعبا على مضيفه فيورنتينا بهدفين لهدف في الأسبوع 35	telegram
49	نجار هولندي يبني سفينة نوح جديدة ستعبر المحيط الأطلسي إلى البرازيل	Twitter
50	صافرات الإنذار تدوي في بلدات إسرائيلية محاذية لقطاع غزة	Twitter
51	إطلاق فاشل ل صاروخ بالستي في كوريا الشمالية	Twitter
52	كريستال بالاس يتأهل لمواجهة مانشستر يونايتد في نهائي كأس إنجلترا بعد فوزه على واتفورد بهدفين لهدف	telegram
53	في نهاية عامها المالي : نجاح بلاي ستيشن 4 يتسبب في زيادة أرباح سوني	Twitter
54	سامسونج تعمل على جهاز جديد للواقع الافتراضي	Twitter
55	مجلس مستوطنات أشكول : لا معلومات بشأن سقوط صواريخ من قطاع غزة بعد دوي صفارات الإنذار	Twitter
56	نتائج سامسونج : جالكسي اس 7 يقود الشركة للنمو مجدداً	Twitter
57	مساعي القاهرة لضبط الدولار	Twitter
58	جماعة أبو سياف تخطف الملاكم العالمي باتشيكوا	Twitter
59	الأهلي يضمن الفوز بلقب الدوري الإماراتي بعد خسارة مطارده العين من الفجيرة 3 - 2	Twitter
60	جنوب السودان يعلن تشكيل حكومة وحدة وطنية جديدة	Twitter
61	مُدافع دورتموند الألماني ماتس هوملز يطلب من ناديه السماح له بالرحيل حتى يتمكن من الانتقال لصفوف بايرن ميونيخ	telegram
62	ليفربول ملزم بدفع تعويض كبير	Twitter
63	تقويم قوقل على أندرويد يجلب ميزة العثور على موعد للإجتماعات	Twitter
64	زيدان يأمل في عودة رونالدو و بنزيمة أمام السيتي	Twitter
65	موقف ميركل من علاقات الناتو و موسكو	Twitter
66	تهدئة بريفي دمشق و اللاذقية تستثني حلب	Twitter
67	اتهام نائبة بريطانية بمعاداة السامية	Twitter
68	حلب_تحترق هاشتاغ يتصدر عالميًا موقع" تويتر " بأكثر من نصف مليون تغريدة تضامنا واستنكارا لجرائم نظام الأسد في حلب	Twitter
69	دل بييرو : سأخبر أحفادي عن ليستر	Twitter
70	دراسة : منتجات الألبان تقلل خطر الإصابة بالسكري	Twitter
71	المسيحيون الشرقيون يحيون الجمعة العظيمة في القدس	Twitter
72	توقف نادر لأشهر ساعة في العالم.. بيج بن تصمت بعد 160 عاماً من الرنين	Facebook

No	Text	Source
73	تعرض قوات الاحتلال لإلقاء زجاجات حارقة قرب جبل المكبر في القدس المحتلة	Facebook
74	مقاتلة روسية ناورت بشكل خطير قرب مقاتلة أميركية فوق بحر البلطيق	Facebook
75	الجينوم و تقنية النانو وتطور الروبوتات هي أبرز ثلاثة مجالات يمكن أن تحدد مستقبل البشرية خلال العقد المقبل	Facebook
76	انصار الصدر يحاولون اقتحام المنطقة الخضراء	Twitter
77	سبب تحطم مروحية روسية في حمص أخطاء بشرية	Twitter
78	الحمض النووي من نقل الجينات البشرية إلى نقل البيانات .. واقع أم مجرد خيال ؟	Facebook
79	كاميرا داعش تسجّل معركة فاشلة ضد البشمركة	telegram
80	مقال بمجلة نيوزويك الأميركية يتساءل عما فعله الرئيس باراك أوباما تجاه جرائم نظام الرئيس السوري بشار الأسد بحق شعبه وقال إن أوباما لم يحرك ساكنا	Facebook
81	الكرملين : بوتين سيلتقي رئيس الوزراء الياباني آبي في مدينة سوتشي في 6 مايو/أيار المقبل	Twitter
82	14 قتيلا بتفجير استهدف زوار شيعة ب العراق	telegram
83	المئات يتظاهرون تضامنا مع حلب في اسطنبول	Facebook
84	وقفة تضامنية مع حلب في أمستردام - هولندا رفضا للمجازر التي يتركبها نظام بشار الأسد وحلفائه الروس	Facebook
85	بوروسيا مونشنغلادباخ يؤجل تتويج بايرن ميونيخ بلقب الدوري الألماني بعد تعادلهما بهدف لمثله في الأسبوع 32	telegram
86	سندرلاند يبقي على آماله في البقاء بفرضه التعادل على مضيفه ستوك سيتي بهدف لمثله في الجولة36	telegram
87	رأسية الويلزي بيل في مرمى ريال سوسييداد التي وضعت ريال مدريد في صدارة الليغا مؤقتا	telegram
88	الدوري الإنجليزي واتفورد يتغلب على ضيفه أستون فيلا بثلاثة أهداف لهدفين في الأسبوع 36 من المسابقة	telegram
89	60مليون هجوم إلكتروني على السعودية في عام واحد	Twitter
90	ليون يحافظ على ترتيبه ب الدوري الفرنسي	Facebook
91	بدء اجتماع الرئاسات الثلاث في العراق بمنزل رئيس الجمهورية فؤاد معصوم	Twitter
92	يوروفيجن تحظر رفع علم فلسطين	Twitter
93	احتفالات ب عيد العمال في إسطنبول	Twitter
94	حتى شرب الماء أوشك أن يكون جريمة!	Twitter
95	أقباط مصر يقيمون قداس عيد الفصح	Twitter
96	باعتباره الخليفة المحتمل للرئيس بوتفليقة .. هل يقوم الوزير السابق شكيب خليل بحملة انتخابية مبكرة	Twitter
97	ساوثهامتون يقسو على مانشستر سيتي ويتغلب عليه بأربعة أهداف لهدفين في الأسبوع 36 و آرسنال ينفرد بالمركز الثالث	telegram
98	لوبان يتنبأ لابنته بالفشل في انتخابات الرئاسة	Twitter
99	سناب شات لمراقبة السائقين المخالفين في السعودية	telegram
100	الدوري الإيطالي لاتسيو يتغلب على ضيفه انتر ميلان بهدفين نظيفين في الأسبوع 36 من المسابقة	telegram

Appendix B: Human Subjects' Opinion

Text_no	Article Title	URL
1	ميسي	https://ar.wikipedia.org/w/index.php?title=%D9%85%D9%8A%D8%B3%D9%8A&redirect=no
1	ريال سبورتينغ خيخون	https://ar.wikipedia.org/wiki/%D8%B1%D9%8A%D8%A7%D9%84_%D8%B3%D8%A8%D9%88%D8%B1%D8%AA%D9%8A%D9%86%D8%BA_%D8%AE%D9%8A%D8%AE%D9%88%D9%86
2	مجلس التعاون الخليجي	https://ar.wikipedia.org/w/index.php?title=%D9%85%D8%AC%D9%84%D8%B3_%D8%A7%D9%84%D8%AA%D8%B9%D8%A7%D9%88%D9%86_%D8%A7%D9%84%D8%AE%D9%84%D9%8A%D8%AC%D9%8A&redirect=no
2	البرلمان العربي	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D8%A8%D8%B1%D9%84%D9%85%D8%A7%D9%86_%D8%A7%D9%84%D8%B9%D8%B1%D8%A8%D9%8A
2	الجامعة العربية	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AC%D8%A7%D9%85%D8%B9%D8%A9_%D8%A7%D9%84%D8%B9%D8%B1%D8%A8%D9%8A%D8%A9&redirect=no
2	حزب الله	https://ar.wikipedia.org/wiki/%D8%AD%D8%B2%D8%A8_%D8%A7%D9%84%D9%84%D9%87
3	الجزيرة الإماراتي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AF%D9%88%D9%84%D8%A7%D8%B1&redirect=no
3	فرانشيسكو توتي	https://ar.wikipedia.org/wiki/%D9%81%D8%B1%D8%A7%D9%86%D8%B4%D9%8A%D8%B3%D9%83%D9%88_%D8%AA%D9%88%D8%AA%D9%8A
4	أوباما	https://ar.wikipedia.org/w/index.php?title=%D8%A3%D9%88%D8%A8%D8%A7%D9%85%D8%A7&redirect=no
4	اللاجئين	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%84%D8%A7%D8%AC%D8%A6%D9%8A%D9%86&redirect=no
5	اللاجئون الفلسطينيون	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%84%D8%A7%D8%AC%D8%A6%D9%88%D9%86_%D8%A7%D9%84%D9%81%D9%84%D8%B3%D8%B7%D9%8A%D9%86%D9%8A%D9%88%D9%86
5	مخيم اليرموك	https://ar.wikipedia.org/wiki/%D9%85%D8%AE%D9%8A%D9%85_%D8%A7%D9%84%D9%8A%D8%B1%D9%85%D9%88%D9%83
5	تنظيم الدولة الإسلامية (داعش)	https://ar.wikipedia.org/wiki/%D8%AA%D9%86%D8%B8%D9%8A%D9%85_%D8%A7%D9%84%D8%AF%D9%88%D9%84%D8%A9_%D8%A7%D9%84%D8%A5%D8%B3%D9%84%D8%A7%D9%85%D9%8A%D8%A9_(%D8%AF%D8%A7%D8%B9%D8%B4)
5	جبهة النصرة	https://ar.wikipedia.org/wiki/%D8%AC%D8%A8%D9%87%D8%A9_%D8%A7%D9%84%D9%86%D8%B5%D8%B1%D8%A9
6	القاهرة	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%82%D8%A7%D9%87%D8%B1%D8%A9
7	ريال مدريد	https://ar.wikipedia.org/wiki/%D8%B1%D9%8A%D8%A7%D9%84_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF
7	نادي خيتافي	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D8%AE%D9%8A%D8%AA%D8%A7%D9%81%D9%8A

Text_no	Article Title	URL
7	نادي برشلونة	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D8%A8%D8%B1%D8%B4%D9%84%D9%88%D9%86%D8%A9
7	أتلتيكو مدريد	https://ar.wikipedia.org/wiki/%D8%A3%D8%AA%D9%84%D8%AA%D9%8A%D9%83%D9%88_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF
7	الليغا	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%84%D9%8A%D8%BA%D8%A7
8	فرانشيسكو توتي	https://ar.wikipedia.org/wiki/%D9%81%D8%B1%D8%A7%D9%86%D8%B4%D9%8A%D8%B3%D9%83%D9%88_%D8%AA%D9%88%D8%AA%D9%8A
8	نادي روما	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D8%B1%D9%88%D9%85%D8%A7
9	نادي برشلونة	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D8%A8%D8%B1%D8%B4%D9%84%D9%88%D9%86%D8%A9
9	أتلتيكو مدريد	https://ar.wikipedia.org/wiki/%D8%A3%D8%AA%D9%84%D8%AA%D9%8A%D9%83%D9%88_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF
10	مانشستر_يونايتد	https://ar.wikipedia.org/wiki/%D9%85%D8%A7%D9%86%D8%B4%D8%B3%D8%AA%D8%B1_%D9%8A%D9%88%D9%86%D8%A7%D9%8A%D8%AA%D8%AF
11	نادي غرناطة	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D8%BA%D8%B1%D9%86%D8%A7%D8%B7%D8%A9
11	نادي ليفانتي	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D9%84%D9%8A%D9%81%D8%A7%D9%86%D8%AA%D9%8A
11	الدوري الإسباني لكرة القدم	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D8%AF%D9%88%D8%B1%D9%8A_%D8%A7%D9%84%D8%A5%D8%B3%D8%A8%D8%A7%D9%86%D9%8A_%D9%84%D9%83%D8%B1%D8%A9_%D8%A7%D9%84%D9%82%D8%AF%D9%85
12	كرستيانو رونالدو	https://ar.wikipedia.org/wiki/%D9%83%D8%B1%D9%8A%D8%B3%D8%AA%D9%8A%D8%A7%D9%86%D9%88_%D8%B1%D9%88%D9%86%D8%A7%D9%84%D8%AF%D9%88
12	زين الدين زيدان	https://ar.wikipedia.org/wiki/%D8%B2%D9%8A%D9%86_%D8%A7%D9%84%D8%AF%D9%8A%D9%86_%D8%B2%D9%8A%D8%AF%D8%A7%D9%86
13	تويتر	https://ar.wikipedia.org/wiki/%D8%AA%D9%88%D9%8A%D8%AA%D8%B1
14	أندرويد	https://ar.wikipedia.org/wiki/%D8%A3%D9%86%D8%AF%D8%B1%D9%88%D9%8A%D8%AF
15	أبل	https://ar.wikipedia.org/wiki/%D8%A3%D8%A8%D9%84
15	آيفون	https://ar.wikipedia.org/w/index.php?title=%D8%A2%D9%8A%D9%81%D9%88%D9%86&redirect=no
16	قوقل	https://ar.wikipedia.org/w/index.php?title=%D9%82%D9%88%D9%82%D9%84&redirect=no
17	رجب طيب أردوغان	https://ar.wikipedia.org/wiki/%D8%B1%D8%AC%D8%A8_%D8%B7%D9%8A%D8%A8_%D8%A3%D8%B1%D8%AF%D9%88%D8%BA%D8%A7%D9%86

Text_no	Article Title	URL
17	الإسلام	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%A5%D8%B3%D9%84%D8%A7%D9%85&redirect=no
18	الجيش العراقي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AC%D9%8A%D8%B4_%D8%A7%D9%84%D8%B9%D8%B1%D8%A7%D9%82%D9%8A&redirect=no
18	الفلوجة	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%81%D9%84%D9%88%D8%AC%D8%A9
18	نينوى	https://ar.wikipedia.org/wiki/%D9%86%D9%8A%D9%86%D9%88%D9%89
19	نظارات غوغل	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%B8%D8%A7%D8%B1%D8%A7%D8%AA_%D8%BA%D9%88%D8%BA%D9%84&redirect=no
19	قوقل	https://ar.wikipedia.org/w/index.php?title=%D9%82%D9%88%D9%82%D9%84&redirect=no
20	نتنياهو	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%AA%D9%86%D9%8A%D8%A7%D9%87%D9%88&redirect=no
20	كتائب القسام	https://ar.wikipedia.org/w/index.php?title=%D9%83%D8%AA%D8%A7%D8%A6%D8%A8_%D8%A7%D9%84%D9%82%D8%B3%D8%A7%D9%85&redirect=no
20	محمد الضيف	https://ar.wikipedia.org/wiki/%D9%85%D8%AD%D9%85%D8%AF_%D8%A7%D9%84%D8%B6%D9%8A%D9%81
21	أوباما	https://ar.wikipedia.org/w/index.php?title=%D8%A3%D9%88%D8%A8%D8%A7%D9%85%D8%A7&redirect=no
21	فلاديمير بوتين	https://ar.wikipedia.org/wiki/%D9%81%D9%84%D8%A7%D8%AF%D9%8A%D9%85%D9%8A%D8%B1_%D8%A8%D9%88%D8%AA%D9%8A%D9%86
21	سوريا	https://ar.wikipedia.org/wiki/%D8%B3%D9%88%D8%B1%D9%8A%D8%A7
22	ميسي	https://ar.wikipedia.org/w/index.php?title=%D9%85%D9%8A%D8%B3%D9%8A&redirect=no
23	تونتنهام هوتسبير	https://ar.wikipedia.org/w/index.php?title=%D8%AA%D9%88%D8%AA%D9%86%D9%87%D8%A7%D9%85_%D9%87%D9%88%D8%AA%D8%B3%D8%A8%D9%8A%D8%B1&redirect=no
23	ليستر سيتي	https://ar.wikipedia.org/wiki/%D9%84%D9%8A%D8%B3%D8%AA%D8%B1_%D8%B3%D9%8A%D8%AA%D9%8A
23	ستوك أون ترينت	https://ar.wikipedia.org/wiki/%D8%B3%D8%AA%D9%88%D9%83_%D8%A3%D9%88%D9%86_%D8%AA%D8%B1%D9%8A%D9%86%D8%AA
24	تنظيم الدولة	https://ar.wikipedia.org/w/index.php?title=%D8%AA%D9%86%D8%B8%D9%8A%D9%85_%D8%A7%D9%84%D8%AF%D9%88%D9%84%D8%A9&redirect=no
24	أميركا	https://ar.wikipedia.org/w/index.php?title=%D8%A3%D9%85%D9%8A%D8%B1%D9%83%D8%A7&redirect=no
24	العراق	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D8%B9%D8%B1%D8%A7%D9%82

Text_no	Article Title	URL
25	اليونسكو	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%8A%D9%88%D9%86%D8%B3%D9%83%D9%88&redirect=no
25	المسجد الأقصى	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%85%D8%B3%D8%AC%D8%AF_%D8%A7%D9%84%D8%A3%D9%82%D8%B5%D9%89
25	جبل الهيكل	https://ar.wikipedia.org/w/index.php?title=%D8%AC%D8%A8%D9%84_%D8%A7%D9%84%D9%87%D9%8A%D9%83%D9%84&redirect=no
26	إسبانيا	https://ar.wikipedia.org/wiki/%D8%A5%D8%B3%D8%A8%D8%A7%D9%86%D9%8A%D8%A7
26	تنظيم الدولة الإسلامية (داعش)	https://ar.wikipedia.org/wiki/%D8%AA%D9%86%D8%B8%D9%8A%D9%85_%D8%A7%D9%84%D8%AF%D9%88%D9%84%D8%A9_%D8%A7%D9%84%D8%A5%D8%B3%D9%84%D8%A7%D9%85%D9%8A%D8%A9_(%D8%AF%D8%A7%D8%B9%D8%B4)
27	مقتدى الصدر	https://ar.wikipedia.org/wiki/%D9%85%D9%82%D8%AA%D8%AF%D9%89_%D8%A7%D9%84%D8%B5%D8%AF%D8%B1
28	ناسا	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%B3%D8%A7
28	تلسكوب هابل الفضائي	https://ar.wikipedia.org/w/index.php?title=%D8%AA%D9%84%D8%B3%D9%83%D9%88%D8%A8_%D9%87%D8%A7%D8%A8%D9%84_%D8%A7%D9%84%D9%81%D8%B6%D8%A7%D8%A6%D9%8A&redirect=no
29	اليمن	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%8A%D9%85%D9%86
30	باراك أوباما	https://ar.wikipedia.org/wiki/%D8%A8%D8%A7%D8%B1%D8%A7%D9%83_%D8%A3%D9%88%D8%A8%D8%A7%D9%85%D8%A7
30	علي خامنئي	https://ar.wikipedia.org/wiki/%D8%B9%D9%84%D9%8A_%D8%AE%D8%A7%D9%85%D9%86%D8%A6%D9%8A
31	فايبر	https://ar.wikipedia.org/wiki/%D9%81%D8%A7%D9%8A%D8%A8%D8%B1
31	واتساب	https://ar.wikipedia.org/w/index.php?title=%D9%88%D8%A7%D8%AA%D8%B3%D8%A7%D8%A8&redirect=no
32	الحوثيون	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AD%D9%88%D8%AB%D9%8A%D9%88%D9%86&redirect=no
33	بغداد	https://ar.wikipedia.org/wiki/%D8%A8%D8%BA%D8%AF%D8%A7%D8%AF
34	أودينيزي	https://ar.wikipedia.org/w/index.php?title=%D8%A3%D9%88%D8%AF%D9%8A%D9%86%D9%8A%D8%B2%D9%8A&redirect=no
34	الإنتر	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%A5%D9%86%D8%AA%D8%B1&redirect=no
35	إسرائيل	https://ar.wikipedia.org/wiki/%D8%A5%D8%B3%D8%B1%D8%A7%D8%A6%D9%8A%D9%84
35	رائد صلاح	https://ar.wikipedia.org/wiki/%D8%B1%D8%A7%D8%A6%D8%AF_%D8%B5%D9%84%D8%A7%D8%AD
36	الحوثي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AD%D9%88%D8%AB%D9%8A&redirect=no
36	تعز	https://ar.wikipedia.org/wiki/%D8%AA%D8%B9%D8%B2

Text_no	Article Title	URL
37	دوري أبطال أوروبا	https://ar.wikipedia.org/wiki/%D8%AF%D9%88%D8%B1%D9%8A_%D8%A3%D8%A8%D8%B7%D8%A7%D9%84_%D8%A3%D9%88%D8%B1%D9%88%D8%A8%D8%A7
37	مانشستر سيتي	https://ar.wikipedia.org/wiki/%D9%85%D8%A7%D9%86%D8%B4%D8%B3%D8%AA%D8%B1_%D8%B3%D9%8A%D8%AA%D9%8A
37	ريال مدريد	https://ar.wikipedia.org/wiki/%D8%B1%D9%8A%D8%A7%D9%84_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF
38	جبل الزاوية	https://ar.wikipedia.org/wiki/%D8%AC%D8%A8%D9%84_%D8%A7%D9%84%D8%B2%D8%A7%D9%88%D9%8A%D8%A9
38	محافظة إدلب	https://ar.wikipedia.org/wiki/%D9%85%D8%AD%D8%A7%D9%81%D8%B8%D8%A9_%D8%A5%D8%AF%D9%84%D8%A8
39	الفلوجة	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%81%D9%84%D9%88%D8%AC%D8%A9
40	وسائل التواصل الاجتماعي	https://ar.wikipedia.org/w/index.php?title=%D9%88%D8%B3%D8%A7%D8%A6%D9%84_%D8%A7%D9%84%D8%AA%D9%88%D8%A7%D8%B5%D9%84_%D8%A7%D9%84%D8%A7%D8%AC%D8%AA%D9%85%D8%A7%D8%B9%D9%8A&redirect=no
40	الولايات المتحدة	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%88%D9%84%D8%A7%D9%8A%D8%A7%D8%AA_%D8%A7%D9%84%D9%85%D8%AA%D8%AD%D8%AF%D8%A9
40	تنظيم الدولة الإسلامية (داعش)	https://ar.wikipedia.org/wiki/%D8%AA%D9%86%D8%B8%D9%8A%D9%85_%D8%A7%D9%84%D8%AF%D9%88%D9%84%D8%A9_%D8%A7%D9%84%D8%A5%D8%B3%D9%84%D8%A7%D9%85%D9%8A%D8%A9_(%D8%AF%D8%A7%D8%B9%D8%B4)
41	البرلمان العراقي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%A8%D8%B1%D9%84%D9%85%D8%A7%D9%86_%D8%A7%D9%84%D8%B9%D8%B1%D8%A7%D9%82%D9%8A&redirect=no
41	إبراهيم الجعفري	https://ar.wikipedia.org/wiki/%D8%A5%D8%A8%D8%B1%D8%A7%D9%87%D9%8A%D9%85_%D8%A7%D9%84%D8%AC%D8%B9%D9%81%D8%B1%D9%8A
42	ريال مدريد	https://ar.wikipedia.org/wiki/%D8%B1%D9%8A%D8%A7%D9%84_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF
42	مانشستر سيتي	https://ar.wikipedia.org/wiki/%D9%85%D8%A7%D9%86%D8%B4%D8%B3%D8%AA%D8%B1_%D8%B3%D9%8A%D8%AA%D9%8A
43	ريال مدريد	https://ar.wikipedia.org/wiki/%D8%B1%D9%8A%D8%A7%D9%84_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF
43	خيسي رودريغيز	https://ar.wikipedia.org/wiki/%D8%AE%D9%8A%D8%B3%D9%8A_%D8%B1%D9%88%D8%AF%D8%B1%D9%8A%D8%BA%D9%8A%D8%B2
44	الأولمبياد	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%A3%D9%88%D9%84%D9%85%D8%A8%D9%8A%D8%A7%D8%AF&redirect=no
44	اليونان	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%8A%D9%88%D9%86%D8%A7%D9%86
45	اليونيسكو	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%8A%D9%88%D9%86%D9%8A%D8%B3%D9%83%D9%88&redirect=no

Text_no	Article Title	URL
45	تدمر	https://ar.wikipedia.org/wiki/%D8%AA%D8%AF%D9%85%D8%B1
46	أتلتيكو مدريد	https://ar.wikipedia.org/wiki/%D8%A3%D8%AA%D9%84%D8%AA%D9%8A%D9%83%D9%88_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF
46	بايرن ميونيخ	https://ar.wikipedia.org/w/index.php?title=%D8%A8%D8%A7%D9%8A%D8%B1%D9%86_%D9%85%D9%8A%D9%88%D9%86%D9%8A%D8%AE&redirect=no
47	النكبة	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%86%D9%83%D8%A8%D8%A9
47	الهولوكوست	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%87%D9%88%D9%84%D9%88%D9%83%D9%88%D8%B3%D8%AA&redirect=no
48	نادي فيورنتينا	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%A7%D8%AF%D9%8A_%D9%81%D9%8A%D9%88%D8%B1%D9%86%D8%AA%D9%8A%D9%86%D8%A7&redirect=no
48	يوفنتوس	https://ar.wikipedia.org/wiki/%D9%8A%D9%88%D9%81%D9%86%D8%AA%D9%88%D8%B3
49	سفينة نوح	https://ar.wikipedia.org/wiki/%D8%B3%D9%81%D9%8A%D9%86%D8%A9_%D9%86%D9%88%D8%AD
49	المحيط الأطلسي	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%85%D8%AD%D9%8A%D8%B7_%D8%A7%D9%84%D8%A3%D8%B7%D9%84%D8%B3%D9%8A
49	البرازيل	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D8%A8%D8%B1%D8%A7%D8%B2%D9%8A%D9%84
50	قطاع غزة	https://ar.wikipedia.org/wiki/%D9%82%D8%B7%D8%A7%D8%B9_%D8%BA%D8%B2%D8%A9
51	صاروخ بالستي	https://ar.wikipedia.org/w/index.php?title=%D8%B5%D8%A7%D8%B1%D9%88%D8%AE_%D8%A8%D8%A7%D9%84%D8%B3%D8%AA%D9%8A&redirect=no
51	كوريا الشمالية	https://ar.wikipedia.org/wiki/%D9%83%D9%88%D8%B1%D9%8A%D8%A7_%D8%A7%D9%84%D8%B4%D9%85%D8%A7%D9%84%D9%8A%D8%A9
52	كريستال بالاس	https://ar.wikipedia.org/wiki/%D9%83%D8%B1%D9%8A%D8%B3%D8%AA%D8%A7%D9%84_%D8%A8%D8%A7%D9%84%D8%A7%D8%B3
52	كأس إنجلترا	https://ar.wikipedia.org/w/index.php?title=%D9%83%D8%A3%D8%B3_%D8%A5%D9%86%D8%AC%D9%84%D8%AA%D8%B1%D8%A7&redirect=no
52	مانشستر يونايتد	https://ar.wikipedia.org/wiki/%D9%85%D8%A7%D9%86%D8%B4%D8%B3%D8%AA%D8%B1_%D9%8A%D9%88%D9%86%D8%A7%D9%8A%D8%AA%D8%AF
52	نادي واتفورد	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D9%88%D8%A7%D8%AA%D9%81%D9%88%D8%B1%D8%AF
53	بلاي ستيشن 4	https://ar.wikipedia.org/wiki/%D8%A8%D9%84%D8%A7%D9%8A_%D8%B3%D8%AA%D9%8A%D8%B4%D9%86_4
53	سوني	https://ar.wikipedia.org/wiki/%D8%B3%D9%88%D9%86%D9%8A

Text_no	Article Title	URL
54	سامسونج	https://ar.wikipedia.org/wiki/%D8%B3%D8%A7%D9%85%D8%B3%D9%88%D9%86%D8%AC
55	قطاع غزة	https://ar.wikipedia.org/wiki/%D9%82%D8%B7%D8%A7%D8%B9_%D8%BA%D8%B2%D8%A9
56	سامسونج	https://ar.wikipedia.org/wiki/%D8%B3%D8%A7%D9%85%D8%B3%D9%88%D9%86%D8%AC
57	القاهرة	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%82%D8%A7%D9%87%D8%B1%D8%A9
57	الدولار الأمريكي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AF%D9%88%D9%84%D8%A7%D8%B1_%D8%A7%D9%84%D8%A3%D9%85%D8%B1%D9%8A%D9%83%D9%8A&redirect=no
58	جماعة أبو سياف	https://ar.wikipedia.org/wiki/%D8%AC%D9%85%D8%A7%D8%B9%D8%A9_%D8%A3%D8%A8%D9%88_%D8%B3%D9%8A%D8%A7%D9%81
59	نادي الفجيرة	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D8%A7%D9%84%D9%81%D8%AC%D9%8A%D8%B1%D8%A9
59	الأهلي الإماراتي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%A3%D9%87%D9%84%D9%8A_%D8%A7%D9%84%D8%A5%D9%85%D8%A7%D8%B1%D8%A7%D8%AA%D9%8A&redirect=no
59	العين الإماراتي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%B9%D9%8A%D9%86_%D8%A7%D9%84%D8%A5%D9%85%D8%A7%D8%B1%D8%A7%D8%AA%D9%8A&redirect=no
60	جنوب السودان	https://ar.wikipedia.org/wiki/%D8%AC%D9%86%D9%88%D8%A8_%D8%A7%D9%84%D8%B3%D9%88%D8%AF%D8%A7%D9%86
61	بايرن ميونيخ	https://ar.wikipedia.org/w/index.php?title=%D8%A8%D8%A7%D9%8A%D8%B1%D9%86_%D9%85%D9%8A%D9%88%D9%86%D9%8A%D8%AE&redirect=no
61	ماتس هوملز	https://ar.wikipedia.org/wiki/%D9%85%D8%A7%D8%AA%D8%B3_%D9%87%D9%88%D9%85%D9%84%D8%B2
61	بوروسيا دورتموند	https://ar.wikipedia.org/wiki/%D8%A8%D9%88%D8%B1%D9%88%D8%B3%D9%8A%D8%A7_%D8%AF%D9%88%D8%B1%D8%AA%D9%85%D9%88%D9%86%D8%AF
62	ليفربول	https://ar.wikipedia.org/wiki/%D9%84%D9%8A%D9%81%D8%B1%D8%A8%D9%88%D9%84
63	قوقل	https://ar.wikipedia.org/w/index.php?title=%D9%82%D9%88%D9%82%D9%84&redirect=no
63	أندرويد	https://ar.wikipedia.org/wiki/%D8%A3%D9%86%D8%AF%D8%B1%D9%88%D9%8A%D8%AF
64	كرستيانو رونالدو	https://ar.wikipedia.org/wiki/%D9%83%D8%B1%D9%8A%D8%B3%D8%AA%D9%8A%D8%A7%D9%86%D9%88_%D8%B1%D9%88%D9%86%D8%A7%D9%84%D8%AF%D9%88
64	زين الدين زيدان	https://ar.wikipedia.org/wiki/%D8%B2%D9%8A%D9%86_%D8%A7%D9%84%D8%AF%D9%8A%D9%86_%D8%B2%D9%8A%D8%AF%D8%A7%D9%86
65	أنغيلا ميركل	https://ar.wikipedia.org/wiki/%D8%A3%D9%86%D8%BA%D9%8A%D9%84%D8%A7_%D9%85%D9%8A%D8%B1%D9%83%D9%84

Text_no	Article Title	URL
65	الناتو (حلف شمال الأطلنطي)	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%86%D8%A7%D8%AA%D9%88_(%D8%AD%D9%84%D9%81_%D8%B4%D9%85%D8%A7%D9%84_%D8%A7%D9%84%D8%A3%D8%B7%D9%84%D9%86%D8%B7%D9%8A)&redirect=no
65	موسكو	https://ar.wikipedia.org/wiki/%D9%85%D9%88%D8%B3%D9%83%D9%88
66	دمشق	https://ar.wikipedia.org/wiki/%D8%AF%D9%85%D8%B4%D9%82
66	اللاذقية	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%84%D8%A7%D8%B0%D9%82%D9%8A%D8%A9
66	حلب	https://ar.wikipedia.org/wiki/%D8%AD%D9%84%D8%A8
67	معاداة السامية	https://ar.wikipedia.org/wiki/%D9%85%D8%B9%D8%A7%D8%AF%D8%A7%D8%A9_%D8%A7%D9%84%D8%B3%D8%A7%D9%85%D9%8A%D8%A9
68	هاشتاغ	https://ar.wikipedia.org/w/index.php?title=%D9%87%D8%A7%D8%B4%D8%AA%D8%A7%D8%BA&redirect=no
68	تغريدة	https://ar.wikipedia.org/w/index.php?title=%D8%AA%D8%BA%D8%B1%D9%8A%D8%AF%D8%A9&redirect=no
68	تويتر	https://ar.wikipedia.org/wiki/%D8%AA%D9%88%D9%8A%D8%AA%D8%B1
68	بشار الأسد	https://ar.wikipedia.org/wiki/%D8%A8%D8%B4%D8%A7%D8%B1_%D8%A7%D9%84%D8%A3%D8%B3%D8%AF
68	حلب	https://ar.wikipedia.org/wiki/%D8%AD%D9%84%D8%A8
69	ألساندرو دل بييرو	https://ar.wikipedia.org/w/index.php?title=%D8%A3%D9%84%D8%B3%D8%A7%D9%86%D8%AF%D8%B1%D9%88_%D8%AF%D9%84_%D8%A8%D9%8A%D9%8A%D8%B1%D9%88&redirect=no
69	ليستر سيتي	https://ar.wikipedia.org/wiki/%D9%84%D9%8A%D8%B3%D8%AA%D8%B1_%D8%B3%D9%8A%D8%AA%D9%8A
70	منتجات_الألبان	https://ar.wikipedia.org/wiki/%D9%85%D9%86%D8%AA%D8%AC%D8%A7%D8%AA_%D8%A7%D9%84%D8%A3%D9%84%D8%A8%D8%A7%D9%86
71	الجمعة العظيمة	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AC%D9%85%D8%B9%D8%A9_%D8%A7%D9%84%D8%B9%D8%B8%D9%8A%D9%85%D8%A9&redirect=no
71	القدس	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%82%D8%AF%D8%B3
72	ساعة بيج بن	https://ar.wikipedia.org/w/index.php?title=%D8%B3%D8%A7%D8%B9%D8%A9_%D8%A8%D9%8A%D8%AC_%D8%A8%D9%86&redirect=no
73	الاحتلال الإسرائيلي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%A7%D8%AD%D8%AA%D9%84%D8%A7%D9%84_%D8%A7%D9%84%D8%A5%D8%B3%D8%B1%D8%A7%D8%A6%D9%8A%D9%84%D9%8A&redirect=no
73	جبل المكبر	https://ar.wikipedia.org/wiki/%D8%AC%D8%A8%D9%84_%D8%A7%D9%84%D9%85%D9%83%D8%A8%D8%B1
73	القدس المحتلة	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%82%D8%AF%D8%B3_%D8%A7%D9%84%D9%85%D8%AD%D8%AA%D9%84%D8%A9&redirect=no

Text_no	Article Title	URL
74	بحر البلطيق	https://ar.wikipedia.org/wiki/%D8%A8%D8%AD%D8%B1_%D8%A7%D9%84%D8%A8%D9%84%D8%B7%D9%8A%D9%82
75	الجينوم	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AC%D9%8A%D9%86%D9%88%D9%85&redirect=no
75	تقنية النانو	https://ar.wikipedia.org/wiki/%D8%AA%D9%82%D9%86%D9%8A%D8%A9_%D8%A7%D9%84%D9%86%D8%A7%D9%86%D9%88
75	الروبوتات	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%B1%D9%88%D8%A8%D9%88%D8%AA%D8%A7%D8%AA&redirect=no
76	المنطقة الخضراء	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%85%D9%86%D8%B7%D9%82%D8%A9_%D8%A7%D9%84%D8%AE%D8%B6%D8%B1%D8%A7%D8%A1
76	مقتدى الصدر	https://ar.wikipedia.org/wiki/%D9%85%D9%82%D8%AA%D8%AF%D9%89_%D8%A7%D9%84%D8%B5%D8%AF%D8%B1
77	حمص	https://ar.wikipedia.org/wiki/%D8%AD%D9%85%D8%B5
78	الحمض النووي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AD%D9%85%D8%B6_%D8%A7%D9%84%D9%86%D9%88%D9%88%D9%8A&redirect=no
78	الجينات	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AC%D9%8A%D9%86%D8%A7%D8%AA&redirect=no
78	نقل البيانات	https://ar.wikipedia.org/wiki/%D9%86%D9%82%D9%84_%D8%A7%D9%84%D8%A8%D9%8A%D8%A7%D9%86%D8%A7%D8%AA
79	تنظيم الدولة الإسلامية (داعش)	https://ar.wikipedia.org/wiki/%D8%AA%D9%86%D8%B8%D9%8A%D9%85_%D8%A7%D9%84%D8%AF%D9%88%D9%84%D8%A9_%D8%A7%D9%84%D8%A5%D8%B3%D9%84%D8%A7%D9%85%D9%8A%D8%A9_(%D8%AF%D8%A7%D8%B9%D8%B4)
79	البشمركة	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%A8%D8%B4%D9%85%D8%B1%D9%83%D8%A9&redirect=no
80	باراك أوباما	https://ar.wikipedia.org/wiki/%D8%A8%D8%A7%D8%B1%D8%A7%D9%83_%D8%A3%D9%88%D8%A8%D8%A7%D9%85%D8%A7
80	نيوزويك	https://ar.wikipedia.org/wiki/%D9%86%D9%8A%D9%88%D8%B2%D9%88%D9%8A%D9%83
80	بشار الأسد	https://ar.wikipedia.org/wiki/%D8%A8%D8%B4%D8%A7%D8%B1_%D8%A7%D9%84%D8%A3%D8%B3%D8%AF
81	الكرملين	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%83%D8%B1%D9%85%D9%84%D9%8A%D9%86&redirect=no
81	فلاديمير بوتين	https://ar.wikipedia.org/wiki/%D9%81%D9%84%D8%A7%D8%AF%D9%8A%D9%85%D9%8A%D8%B1_%D8%A8%D9%88%D8%AA%D9%8A%D9%86
81	شينزو آبي	https://ar.wikipedia.org/wiki/%D8%B4%D9%8A%D9%86%D8%B2%D9%88_%D8%A2%D8%A8%D9%8A
81	سوتشي	https://ar.wikipedia.org/wiki/%D8%B3%D9%88%D8%AA%D8%B4%D9%8A
82	الشيعة	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D8%B4%D9%8A%D8%B9%D8%A9
82	العراق	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D8%B9%D8%B1%D8%A7%D9%82

Text_no	Article Title	URL
83	اسطنبول	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D8%B3%D8%B7%D9%86%D8%A8%D9%88%D9%84&redirect=no
83	حلب	https://ar.wikipedia.org/wiki/%D8%AD%D9%84%D8%A8
84	بشار الأسد	https://ar.wikipedia.org/wiki/%D8%A8%D8%B4%D8%A7%D8%B1_%D8%A7%D9%84%D8%A3%D8%B3%D8%AF
84	حلب	https://ar.wikipedia.org/wiki/%D8%AD%D9%84%D8%A8
84	أمستردام	https://ar.wikipedia.org/wiki/%D8%A3%D9%85%D8%B3%D8%AA%D8%B1%D8%AF%D8%A7%D9%85
84	هولندا	https://ar.wikipedia.org/wiki/%D9%87%D9%88%D9%84%D9%86%D8%AF%D8%A7
85	نادي بوروسيا مونشنغلادباخ	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%A7%D8%AF%D9%8A_%D8%A8%D9%88%D8%B1%D9%88%D8%B3%D9%8A%D8%A7_%D9%85%D9%88%D9%86%D8%B4%D9%86%D8%BA%D9%84%D8%A7%D8%AF%D8%A8%D8%A7%D8%AE&redirect=no
85	الدوري الألماني الدرجة الأولى	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AF%D9%88%D8%B1%D9%8A_%D8%A7%D9%84%D8%A3%D9%84%D9%85%D8%A7%D9%86%D9%8A_%D8%A7%D9%84%D8%AF%D8%B1%D8%AC%D8%A9_%D8%A7%D9%84%D8%A3%D9%88%D9%84%D9%89&redirect=no
85	بايرن ميونيخ	https://ar.wikipedia.org/w/index.php?title=%D8%A8%D8%A7%D9%8A%D8%B1%D9%86_%D9%85%D9%8A%D9%88%D9%86%D9%8A%D8%AE&redirect=no
86	ستوك سيتي	https://ar.wikipedia.org/wiki/%D8%B3%D8%AA%D9%88%D9%83_%D8%B3%D9%8A%D8%AA%D9%8A
86	نادي سندرلاند	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D8%B3%D9%86%D8%AF%D8%B1%D9%84%D8%A7%D9%86%D8%AF
87	غاريث بيل	https://ar.wikipedia.org/wiki/%D8%BA%D8%A7%D8%B1%D9%8A%D8%AB_%D8%A8%D9%8A%D9%84
87	ريال سوسييداد	https://ar.wikipedia.org/w/index.php?title=%D8%B1%D9%8A%D8%A7%D9%84_%D8%B3%D9%88%D8%B3%D9%8A%D9%8A%D8%AF%D8%A7%D8%AF&redirect=no
87	الليغا	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%84%D9%8A%D8%BA%D8%A7&redirect=no
87	نادي ريال مدريد	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%A7%D8%AF%D9%8A_%D8%B1%D9%8A%D8%A7%D9%84_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF&redirect=no
88	الدوري الإنجليزي الممتاز	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D8%AF%D9%88%D8%B1%D9%8A_%D8%A7%D9%84%D8%A5%D9%86%D8%AC%D9%84%D9%8A%D8%B2%D9%8A_%D8%A7%D9%84%D9%85%D9%85%D8%AA%D8%A7%D8%B2
88	نادي أستون فيلا	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%A7%D8%AF%D9%8A_%D8%A3%D8%B3%D8%AA%D9%88%D9%86_%D9%81%D9%8A%D9%84%D8%A7&redirect=no
88	نادي واتفورد	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D9%88%D8%A7%D8%AA%D9%81%D9%88%D8%B1%D8%AF

Text_no	Article Title	URL
89	السعودية	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D8%B3%D8%B9%D9%88%D8%AF%D9%8A%D8%A9
90	أولمبيك ليون	https://ar.wikipedia.org/wiki/%D8%A3%D9%88%D9%84%D9%85%D8%A8%D9%8A%D9%83_%D9%84%D9%8A%D9%88%D9%86
90	الدوري الفرنسي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AF%D9%88%D8%B1%D9%8A_%D8%A7%D9%84%D9%81%D8%B1%D9%86%D8%B3%D9%8A&redirect=no
91	فؤاد معصوم	https://ar.wikipedia.org/wiki/%D9%81%D8%A4%D8%A7%D8%AF_%D9%85%D8%B9%D8%B5%D9%88%D9%85
91	العراق	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D8%B9%D8%B1%D8%A7%D9%82
92	يوروفيجن	https://ar.wikipedia.org/w/index.php?title=%D9%8A%D9%88%D8%B1%D9%88%D9%81%D9%8A%D8%AC%D9%86&redirect=no
92	علم فلسطين	https://ar.wikipedia.org/wiki/%D8%B9%D9%84%D9%85_%D9%81%D9%84%D8%B3%D8%B7%D9%8A%D9%86
93	إسطنبول	https://ar.wikipedia.org/wiki/%D8%A5%D8%B3%D8%B7%D9%86%D8%A8%D9%88%D9%84
93	عيد العمال العالمي	https://ar.wikipedia.org/w/index.php?title=%D8%B9%D9%8A%D8%AF_%D8%A7%D9%84%D8%B9%D9%85%D8%A7%D9%84_%D8%A7%D9%84%D8%B9%D8%A7%D9%84%D9%85%D9%8A&redirect=no
94	الماء	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%85%D8%A7%D8%A1&redirect=no
95	اقباط	https://ar.wikipedia.org/wiki/%D8%A3%D9%82%D8%A8%D8%A7%D8%B7
95	عيد الفصح	https://ar.wikipedia.org/w/index.php?title=%D8%B9%D9%8A%D8%AF_%D8%A7%D9%84%D9%81%D8%B5%D8%AD&redirect=no
96	عبد العزيز بوتفليقة	https://ar.wikipedia.org/wiki/%D8%B9%D8%A8%D8%AF_%D8%A7%D9%84%D8%B9%D8%B2%D9%8A%D8%B2_%D8%A8%D9%88%D8%AA%D9%81%D9%84%D9%8A%D9%82%D8%A9
96	شكيب خليل	https://ar.wikipedia.org/wiki/%D8%B4%D9%83%D9%8A%D8%A8_%D8%AE%D9%84%D9%8A%D9%84
97	نادي مانشستر سيتي	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%A7%D8%AF%D9%8A_%D9%85%D8%A7%D9%86%D8%B4%D8%B3%D8%AA%D8%B1_%D8%B3%D9%8A%D8%AA%D9%8A&redirect=no
97	آرسنال	https://ar.wikipedia.org/w/index.php?title=%D8%A2%D8%B1%D8%B3%D9%86%D8%A7%D9%84&redirect=no
98	جان ماري لوبان	https://ar.wikipedia.org/wiki/%D8%AC%D8%A7%D9%86_%D9%85%D8%A7%D8%B1%D9%8A_%D9%84%D9%88%D8%A8%D8%A7%D9%86
99	سناب شات	https://ar.wikipedia.org/wiki/%D8%B3%D9%86%D8%A7%D8%A8_%D8%B4%D8%A7%D8%AA
99	المملكة العربية السعودية	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%85%D9%85%D9%84%D9%83%D8%A9_%D8%A7%D9%84%D8%B9%D8%B1%D8%A8%D9%8A%D8%A9_%D8%A7%D9%84%D8%B3%D8%B9%D9%88%D8%AF%D9%8A%D8%A9&redirect=no

Text_no	Article Title	URL
100	نادي لاتسيو	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D9%84%D8%A7%D8%AA%D8%B3%D9%8A%D9%88
100	الدوري الإيطالي الممتاز	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AF%D9%88%D8%B1%D9%8A_%D8%A7%D9%84%D8%A5%D9%8A%D8%B7%D8%A7%D9%84%D9%8A_%D8%A7%D9%84%D9%85%D9%85%D8%AA%D8%A7%D8%B2&redirect=no
100	انتر ميلان	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%86%D8%AA%D8%B1_%D9%85%D9%8A%D9%84%D8%A7%D9%86&redirect=no

Appendix B 1: System Results

Text_no	Article Title	URL
1	ميسي	https://ar.wikipedia.org/w/index.php?title=%D9%85%D9%8A%D8%B3%D9%8A&redirect=no
1	سبورتنغ خيخون	https://ar.wikipedia.org/w/index.php?title=%D8%B3%D8%A8%D9%88%D8%B1%D8%AA%D9%8A%D9%86%D8%BA_%D8%AE%D9%8A%D8%AE%D9%88%D9%86&redirect=no
2	مجلس التعاون الخليجي	https://ar.wikipedia.org/w/index.php?title=%D9%85%D8%AC%D9%84%D8%B3_%D8%A7%D9%84%D8%AA%D8%B9%D8%A7%D9%88%D9%86_%D8%A7%D9%84%D8%AE%D9%84%D9%8A%D8%AC%D9%8A&redirect=no
2	كتائب حزب الله	https://ar.wikipedia.org/wiki/%D9%83%D8%AA%D8%A7%D8%A6%D8%A8_%D8%AD%D8%B2%D8%A8_%D8%A7%D9%84%D9%84%D9%87
2	جماعة الجهاد الإسلامي	https://ar.wikipedia.org/wiki/%D8%AC%D9%85%D8%A7%D8%B9%D8%A9_%D8%A7%D9%84%D8%AC%D9%87%D8%A7%D8%AF_%D8%A7%D9%84%D8%A5%D8%B3%D9%84%D8%A7%D9%85%D9%8A
2	الجامعة العربية	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AC%D8%A7%D9%85%D8%B9%D8%A9_%D8%A7%D9%84%D8%B9%D8%B1%D8%A8%D9%8A%D8%A9&redirect=no
2	منظمة إرهابية	https://ar.wikipedia.org/w/index.php?title=%D9%85%D9%86%D8%B8%D9%85%D8%A9_%D8%A5%D8%B1%D9%87%D8%A7%D8%A8%D9%8A%D8%A9&redirect=no
3	فرانشيسكو توتي	https://ar.wikipedia.org/wiki/%D9%81%D8%B1%D8%A7%D9%86%D8%B4%D9%8A%D8%B3%D9%83%D9%88_%D8%AA%D9%88%D8%AA%D9%8A
3	الجزيرة الإماراتي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AC%D8%B2%D9%8A%D8%B1%D8%A9_%D8%A7%D9%84%D8%A5%D9%85%D8%A7%D8%B1%D8%A7%D8%AA%D9%8A&redirect=no
4	أوباما	https://ar.wikipedia.org/w/index.php?title=%D8%A3%D9%88%D8%A8%D8%A7%D9%85%D8%A7&redirect=no
4	اللاجئين	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%84%D8%A7%D8%AC%D8%A6%D9%8A%D9%86&redirect=no
4	ندين تحسين بيك	https://ar.wikipedia.org/wiki/%D9%86%D8%AF%D9%8A%D9%86_%D8%AA%D8%AD%D8%B3%D9%8A%D9%86_%D8%A8%D9%83
5	اللاجئون الفلسطينيون في سورية	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%84%D8%A7%D8%AC%D8%A6%D9%88%D9%86_%D8%A7%D9%84%D9%81%D9%84%D8%B3%D8%B7%D9%8A%D9%86%D9%8A%D9%88%D9%86_%D9%81%D9%8A_%D8%B3%D9%88%D8%B1%D9%8A%D8%A9&redirect=no

Text_no	Article Title	URL
5	حصار اللاذقية	https://ar.wikipedia.org/wiki/%D8%AD%D8%B5%D8%A7%D8%B1_%D8%A7%D9%84%D9%84%D8%A7%D8%B0%D9%82%D9%8A%D8%A9
5	تنظيم الدولة الإسلامية (داعش)	https://ar.wikipedia.org/wiki/%D8%AA%D9%86%D8%B8%D9%8A%D9%85_%D8%A7%D9%84%D8%AF%D9%88%D9%84%D8%A9_%D8%A7%D9%84%D8%A5%D8%B3%D9%84%D8%A7%D9%85%D9%8A%D8%A9_(%D8%AF%D8%A7%D8%B9%D8%B4)
5	جبهة النصرة	https://ar.wikipedia.org/wiki/%D8%AC%D8%A8%D9%87%D8%A9_%D8%A7%D9%84%D9%86%D8%B5%D8%B1%D8%A9
5	معركة مخيم اليرموك (2015)	https://ar.wikipedia.org/wiki/%D9%85%D8%B9%D8%B1%D9%83%D8%A9_%D9%85%D8%AE%D9%8A%D9%85_%D8%A7%D9%84%D9%8A%D8%B1%D9%85%D9%88%D9%83_(2015)
6	القاهرة	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%82%D8%A7%D9%87%D8%B1%D8%A9
7	نادي خيتافي	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D8%AE%D9%8A%D8%AA%D8%A7%D9%81%D9%8A
7	تاريخ ريال مدريد	https://ar.wikipedia.org/wiki/%D8%AA%D8%A7%D8%B1%D9%8A%D8%AE_%D8%B1%D9%8A%D8%A7%D9%84_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF
7	إسبانيول برشلونة	https://ar.wikipedia.org/wiki/%D8%A5%D8%B3%D8%A8%D8%A7%D9%86%D9%8A%D9%88%D9%84_%D8%A8%D8%B1%D8%B4%D9%84%D9%88%D9%86%D8%A9
7	الليغا	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%84%D9%8A%D8%BA%D8%A7
9	نادي برشلونة	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D8%A8%D8%B1%D8%B4%D9%84%D9%88%D9%86%D8%A9
10	مانشستر_يونايتد	https://ar.wikipedia.org/wiki/%D9%85%D8%A7%D9%86%D8%B4%D8%B3%D8%AA%D8%B1_%D9%8A%D9%88%D9%86%D8%A7%D9%8A%D8%AA%D8%AF
11	نادي غرناطة	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D8%BA%D8%B1%D9%86%D8%A7%D8%B7%D8%A9
12	كرستيانو رونالدو	https://ar.wikipedia.org/wiki/%D9%83%D8%B1%D9%8A%D8%B3%D8%AA%D9%8A%D8%A7%D9%86%D9%88_%D8%B1%D9%88%D9%86%D8%A7%D9%84%D8%AF%D9%88
12	زين الدين زيدان	https://ar.wikipedia.org/wiki/%D8%B2%D9%8A%D9%86_%D8%A7%D9%84%D8%AF%D9%8A%D9%86_%D8%B2%D9%8A%D8%AF%D8%A7%D9%86
13	تفجيرات الرياض	https://ar.wikipedia.org/wiki/%D8%AA%D9%81%D8%AC%D9%8A%D8%B1%D8%A7%D8%AA_%D8%A7%D9%84%D8%B1%D9%8A%D8%A7%D8%B6
13	خمسة	https://ar.wikipedia.org/wiki/%D8%AE%D9%85%D8%B3%D8%A9
13	تويتر	https://ar.wikipedia.org/wiki/%D8%AA%D9%88%D9%8A%D8%AA%D8%B1
14	أندرويد	https://ar.wikipedia.org/wiki/%D8%A3%D9%86%D8%AF%D8%B1%D9%88%D9%8A%D8%AF
15	الجرائم	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AC%D8%B1%D8%A7%D8%A6%D9%85&redirect=no
15	ثقافة أمريكية جنوبية	https://ar.wikipedia.org/wiki/%D8%AB%D9%82%D8%A7%D9%81%D8%A9_%D8%A3%D9%85%D8%B1%D9%8A%D9%83%D9%8

Text_no	Article Title	URL
		A%D8%A9_%D8%AC%D9%86%D9%88%D8%A8%D9%8A%D8%A9
15	آيفون	https://ar.wikipedia.org/w/index.php?title=%D8%A2%D9%8A%D9%81%D9%88%D9%86&redirect=no
15	مطورو آبل	https://ar.wikipedia.org/wiki/%D9%85%D8%B7%D9%88%D8%B1%D9%88_%D8%A2%D8%A8%D9%84
16	قوقل	https://ar.wikipedia.org/w/index.php?title=%D9%82%D9%88%D9%82%D9%84&redirect=no
17	قائمة المنظمات المصنفة كإرهابية	https://ar.wikipedia.org/w/index.php?title=%D9%82%D8%A7%D8%A6%D9%85%D8%A9_%D8%A7%D9%84%D9%85%D9%86%D8%B8%D9%85%D8%A7%D8%AA_%D8%A7%D9%84%D9%85%D8%B5%D9%86%D9%81%D8%A9_%D9%83%D8%A5%D8%B1%D9%87%D8%A7%D8%A8%D9%8A%D8%A9&redirect=no
18	الجيش العراقي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AC%D9%8A%D8%B4_%D8%A7%D9%84%D8%B9%D8%B1%D8%A7%D9%82%D9%8A&redirect=no
18	الفلوجة	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%81%D9%84%D9%88%D8%AC%D8%A9
18	سهل نينوى	https://ar.wikipedia.org/wiki/%D8%B3%D9%87%D9%84_%D9%86%D9%8A%D9%86%D9%88%D9%89
19	نظارات غوغل	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%B8%D8%A7%D8%B1%D8%A7%D8%AA_%D8%BA%D9%88%D8%BA%D9%84&redirect=no
19	قوقل	https://ar.wikipedia.org/w/index.php?title=%D9%82%D9%88%D9%82%D9%84&redirect=no
20	حساب	https://ar.wikipedia.org/wiki/%D8%AD%D8%B3%D8%A7%D8%A8
20	المعارضة	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%85%D8%B9%D8%A7%D8%B1%D8%B6%D8%A9&redirect=no
20	نتنياهو	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%AA%D9%86%D9%8A%D8%A7%D9%87%D9%88&redirect=no
20	كتائب القسام	https://ar.wikipedia.org/w/index.php?title=%D9%83%D8%AA%D8%A7%D8%A6%D8%A8_%D8%A7%D9%84%D9%82%D8%B3%D8%A7%D9%85&redirect=no
20	محمد الضيف	https://ar.wikipedia.org/wiki/%D9%85%D8%AD%D9%85%D8%AF_%D8%A7%D9%84%D8%B6%D9%8A%D9%81
20	زعيم	https://ar.wikipedia.org/w/index.php?title=%D8%B2%D8%B9%D9%8A%D9%85&redirect=no
20	الجدار الإسرائيلي العازل	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AC%D8%AF%D8%A7%D8%B1_%D8%A7%D9%84%D8%A5%D8%B3%D8%B1%D8%A7%D8%A6%D9%8A%D9%84%D9%8A_%D8%A7%D9%84%D8%B9%D8%A7%D8%B2%D9%84&redirect=no
21	أوباما	https://ar.wikipedia.org/w/index.php?title=%D8%A3%D9%88%D8%A8%D8%A7%D9%85%D8%A7&redirect=no
22	ميسي	https://ar.wikipedia.org/w/index.php?title=%D9%85%D9%8A%D8%B3%D9%8A&redirect=no
23	تونتنهام هوتسبير	https://ar.wikipedia.org/w/index.php?title=%D8%AA%D9%88%D8%AA%D9%86%D9%87%D8%A7%D9%85_%D9%87%D9%88%D8%AA%D8%B3%D8%A8%D9%8A%D8%B1&redirect=no
23	ليستر سيتي	https://ar.wikipedia.org/wiki/%D9%84%D9%8A%D8%B3%D8%AA%D8%B1_%D8%B3%D9%8A%D8%AA%D9%8A

Text_no	Article Title	URL
23	ستوك أون ترينت	https://ar.wikipedia.org/wiki/%D8%B3%D8%AA%D9%88%D9%83_%D8%A3%D9%88%D9%86_%D8%AA%D8%B1%D9%8A%D9%86%D8%AA
24	تنظيم الدولة	https://ar.wikipedia.org/w/index.php?title=%D8%AA%D9%86%D8%B8%D9%8A%D9%85_%D8%A7%D9%84%D8%AF%D9%88%D9%84%D8%A9&redirect=no
24	القاعدة في العراق	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%82%D8%A7%D8%B9%D8%AF%D8%A9_%D9%81%D9%8A_%D8%A7%D9%84%D8%B9%D8%B1%D8%A7%D9%82
25	قائمة مساطب المسجد الأقصى	https://ar.wikipedia.org/wiki/%D9%82%D8%A7%D8%A6%D9%85%D8%A9_%D9%85%D8%B3%D8%A7%D8%B7%D8%A8_%D8%A7%D9%84%D9%85%D8%B3%D8%AC%D8%AF_%D8%A7%D9%84%D8%A3%D9%82%D8%B5%D9%89
25	اليونسكو	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%8A%D9%88%D9%86%D8%B3%D9%83%D9%88&redirect=no
25	أمناء جبل الهيكل	https://ar.wikipedia.org/wiki/%D8%A3%D9%85%D9%86%D8%A7%D8%A1_%D8%AC%D8%A8%D9%84_%D8%A7%D9%84%D9%87%D9%8A%D9%83%D9%84
26	إسبانيا	https://ar.wikipedia.org/wiki/%D8%A5%D8%B3%D8%A8%D8%A7%D9%86%D9%8A%D8%A7
27	مقتدى الصدر	https://ar.wikipedia.org/wiki/%D9%85%D9%82%D8%AA%D8%AF%D9%89_%D8%A7%D9%84%D8%B5%D8%AF%D8%B1
27	الاثنين	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D8%A7%D8%AB%D9%86%D9%8A%D9%86
28	ناسا	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%B3%D8%A7
28	تلسكوب هابل الفضائي	https://ar.wikipedia.org/w/index.php?title=%D8%AA%D9%84%D8%B3%D9%83%D9%88%D8%A8_%D9%87%D8%A7%D8%A8%D9%84_%D8%A7%D9%84%D9%81%D8%B6%D8%A7%D8%A6%D9%8A&redirect=no
28	منصة إطلاق	https://ar.wikipedia.org/wiki/%D9%85%D9%86%D8%B5%D8%A9_%D8%A5%D8%B7%D9%84%D8%A7%D9%82
29	اليمن	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%8A%D9%85%D9%86
30	باراك أوباما	https://ar.wikipedia.org/wiki/%D8%A8%D8%A7%D8%B1%D8%A7%D9%83_%D8%A3%D9%88%D8%A8%D8%A7%D9%85%D8%A7
30	علي خامنئي	https://ar.wikipedia.org/wiki/%D8%B9%D9%84%D9%8A_%D8%AE%D8%A7%D9%85%D9%86%D8%A6%D9%8A
31	فايبر	https://ar.wikipedia.org/wiki/%D9%81%D8%A7%D9%8A%D8%A8%D8%B1
31	واتساب	https://ar.wikipedia.org/w/index.php?title=%D9%88%D8%A7%D8%AA%D8%B3%D8%A7%D8%A8&redirect=no
32	الحوثيون	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AD%D9%88%D8%AB%D9%8A%D9%88%D9%86&redirect=no
33	بغداد	https://ar.wikipedia.org/wiki/%D8%A8%D8%BA%D8%AF%D8%A7%D8%AF
34	أودينيزي	https://ar.wikipedia.org/w/index.php?title=%D8%A3%D9%88%D8%AF%D9%8A%D9%86%D9%8A%D8%B2%D9%8A&redirect=no
34	الإنتر	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%A5%D9%86%D8%AA%D8%B1&redirect=no

Text_no	Article Title	URL
35	التعليم في إسرائيل	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D8%AA%D8%B9%D9%84%D9%8A%D9%85_%D9%81%D9%8A_%D8%A5%D8%B3%D8%B1%D8%A7%D8%A6%D9%8A%D9%84
35	الشيخ رائد صلاح	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%B4%D9%8A%D8%AE_%D8%B1%D8%A7%D8%A6%D8%AF_%D8%B5%D9%84%D8%A7%D8%AD&redirect=no
36	محمد علي الحوثي	https://ar.wikipedia.org/wiki/%D9%85%D8%AD%D9%85%D8%AF_%D8%B9%D9%84%D9%8A_%D8%A7%D9%84%D8%AD%D9%88%D8%AB%D9%8A
36	معركة تعز (2015)	https://ar.wikipedia.org/w/index.php?title=%D9%85%D8%B9%D8%B1%D9%83%D8%A9_%D8%AA%D8%B9%D8%B2_(2015)&redirect=no
37	دوري أبطال أوروبا 08-2007	https://ar.wikipedia.org/wiki/%D8%AF%D9%88%D8%B1%D9%8A_%D8%A3%D8%A8%D8%B7%D8%A7%D9%84_%D8%A3%D9%88%D8%B1%D9%88%D8%A8%D8%A7_2007%E2%80%9308
37	مانشستر سيتي	https://ar.wikipedia.org/wiki/%D9%85%D8%A7%D9%86%D8%B4%D8%B3%D8%AA%D8%B1_%D8%B3%D9%8A%D8%AA%D9%8A
37	إحصائيات نادي ريال مدريد	https://ar.wikipedia.org/wiki/%D8%A5%D8%AD%D8%B5%D8%A7%D8%A6%D9%8A%D8%A7%D8%AA_%D8%B1%D9%8A%D8%A7%D9%84_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF
38	معركة إدلب (2012)	https://ar.wikipedia.org/w/index.php?title=%D9%85%D8%B9%D8%B1%D9%83%D8%A9_%D8%A5%D8%AF%D9%84%D8%A8_(2012)&redirect=no
38	مجزرة جبل الزاوية (ديسمبر 2011)	https://ar.wikipedia.org/w/index.php?title=%D9%85%D8%AC%D8%B2%D8%B1%D8%A9_%D8%AC%D8%A8%D9%84_%D8%A7%D9%84%D8%B2%D8%A7%D9%88%D9%8A%D8%A9_(%D8%AF%D9%8A%D8%B3%D9%85%D8%A8%D8%B1_2011)&redirect=no
39	أحداث الفلوجة	https://ar.wikipedia.org/wiki/%D8%A3%D8%AD%D8%AF%D8%A7%D8%AB_%D8%A7%D9%84%D9%81%D9%84%D9%88%D8%AC%D8%A9
39	المؤتمر الوطني العراقي	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%85%D8%A4%D8%AA%D9%85%D8%B1_%D8%A7%D9%84%D9%88%D8%B7%D9%86%D9%8A_%D8%A7%D9%84%D8%B9%D8%B1%D8%A7%D9%82%D9%8A
40	حوسبة فائقة	https://ar.wikipedia.org/wiki/%D8%AD%D9%88%D8%B3%D8%A8%D8%A9_%D9%81%D8%A7%D8%A6%D9%82%D8%A9
40	وسائل التواصل الاجتماعي	https://ar.wikipedia.org/w/index.php?title=%D9%88%D8%B3%D8%A7%D8%A6%D9%84_%D8%A7%D9%84%D8%AA%D9%88%D8%A7%D8%B5%D9%84_%D8%A7%D9%84%D8%A7%D8%AC%D8%AA%D9%85%D8%A7%D8%B9%D9%8A&redirect=no
40	تنظيم الدولة	https://ar.wikipedia.org/w/index.php?title=%D8%AA%D9%86%D8%B8%D9%8A%D9%85_%D8%A7%D9%84%D8%AF%D9%88%D9%84%D8%A9&redirect=no
40	الولايات المتحدة الكولومبية	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%88%D9%84%D8%A7%D9%8A%D8%A7%D8%AA_%D8%A7%D9%84%D9%85%D8%AA%D8%AD%D8%AF%D8%A9_%D8%A7%D9%84%D9%83%D9%88%D9%84%D9%88%D9%85%D8%A8%D9%8A%D8%A9

Text_no	Article Title	URL
40	جبهة وطنية (كولومبيا)	https://ar.wikipedia.org/wiki/%D8%AC%D8%A8%D9%87%D8%A9_%D9%88%D8%B7%D9%86%D9%8A%D8%A9_(%D9%83%D9%88%D9%84%D9%88%D9%85%D8%A8%D9%8A%D8%A7)
41	إبراهيم الجعفري	https://ar.wikipedia.org/wiki/%D8%A5%D8%A8%D8%B1%D8%A7%D9%87%D9%8A%D9%85_%D8%A7%D9%84%D8%AC%D8%B9%D9%81%D8%B1%D9%8A
41	استقالة	https://ar.wikipedia.org/wiki/%D8%A7%D8%B3%D8%AA%D9%82%D8%A7%D9%84%D8%A9
41	وزير الخارجية الأمريكي	https://ar.wikipedia.org/wiki/%D9%88%D8%B2%D9%8A%D8%B1_%D8%A7%D9%84%D8%AE%D8%A7%D8%B1%D8%AC%D9%8A%D8%A9_%D8%A7%D9%84%D8%A3%D9%85%D8%B1%D9%8A%D9%83%D9%8A
41	البرلمان العراقي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%A8%D8%B1%D9%84%D9%85%D8%A7%D9%86_%D8%A7%D9%84%D8%B9%D8%B1%D8%A7%D9%82%D9%8A&redirect=no
42	سيد أولمبيا	https://ar.wikipedia.org/wiki/%D8%B3%D9%8A%D8%AF_%D8%A3%D9%88%D9%84%D9%8A%D9%85%D8%A8%D9%8A%D8%A7
42	الموقف الأخير	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%85%D9%88%D9%82%D9%81_%D8%A7%D9%84%D8%A3%D8%AE%D9%8A%D8%B1
42	نادي ريال مدريد	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%A7%D8%AF%D9%8A_%D8%B1%D9%8A%D8%A7%D9%84_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF&redirect=no
42	مانشستر سيتي	https://ar.wikipedia.org/wiki/%D9%85%D8%A7%D9%86%D8%B4%D8%B3%D8%AA%D8%B1_%D8%B3%D9%8A%D8%AA%D9%8A
43	أول موفي	https://ar.wikipedia.org/wiki/%D8%A3%D9%88%D9%84_%D9%85%D9%88%D9%81%D9%8A
43	ريال مدريد	https://ar.wikipedia.org/wiki/%D8%B1%D9%8A%D8%A7%D9%84_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF
44	شتات سوري	https://ar.wikipedia.org/wiki/%D8%B4%D8%AA%D8%A7%D8%AA_%D8%B3%D9%88%D8%B1%D9%8A
44	لاجئ	https://ar.wikipedia.org/wiki/%D9%84%D8%A7%D8%AC%D8%A6
44	عرب اليونان	https://ar.wikipedia.org/wiki/%D8%B9%D8%B1%D8%A8_%D8%A7%D9%84%D9%8A%D9%88%D9%86%D8%A7%D9%86
45	مملكة تدمر	https://ar.wikipedia.org/wiki/%D9%85%D9%85%D9%84%D9%83%D8%A9_%D8%AA%D8%AF%D9%85%D8%B1
45	اليونيسكو	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%8A%D9%88%D9%86%D9%8A%D8%B3%D9%83%D9%88&redirect=no
46	دوري أبطال أوروبا 13-2012	https://ar.wikipedia.org/wiki/%D8%AF%D9%88%D8%B1%D9%8A_%D8%A3%D8%A8%D8%B7%D8%A7%D9%84_%D8%A3%D9%88%D8%B1%D9%88%D8%A8%D8%A7_2012-13
46	أتلتيكو مدريد	https://ar.wikipedia.org/wiki/%D8%A3%D8%AA%D9%84%D8%AA%D9%8A%D9%83%D9%88_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF
46	بايرن ميونيخ	https://ar.wikipedia.org/wiki/%D8%A8%D8%A7%D9%8A%D8%B1%D9%86_%D9%85%D9%8A%D9%88%D9%86%D8%AE

Text_no	Article Title	URL
47	ذكرى النكبة	https://ar.wikipedia.org/wiki/%D8%B0%D9%83%D8%B1%D9%89_%D8%A7%D9%84%D9%86%D9%83%D8%A8%D8%A9
47	الهولوكوست	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%87%D9%88%D9%84%D9%88%D9%83%D9%88%D8%B3%D8%AA&redirect=no
48	نادي فيورنتينا	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%A7%D8%AF%D9%8A_%D9%81%D9%8A%D9%88%D8%B1%D9%86%D8%AA%D9%8A%D9%86%D8%A7&redirect=no
48	يوفنتوس	https://ar.wikipedia.org/wiki/%D9%8A%D9%88%D9%81%D9%86%D8%AA%D9%88%D8%B3
49	سفينة نوح	https://ar.wikipedia.org/wiki/%D8%B3%D9%81%D9%8A%D9%86%D8%A9_%D9%86%D9%88%D8%AD
49	أدب هولندي	https://ar.wikipedia.org/w/index.php?title=%D8%A3%D8%AF%D8%A8_%D9%87%D9%88%D9%84%D9%86%D8%AF%D9%8A&redirect=no
49	تاريخ البرازيل	https://ar.wikipedia.org/wiki/%D8%AA%D8%A7%D8%B1%D9%8A%D8%AE_%D8%A7%D9%84%D8%A8%D8%B1%D8%A7%D8%B2%D9%8A%D9%84
49	نجار	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%AC%D8%A7%D8%B1&redirect=no
49	جديدة	https://ar.wikipedia.org/w/index.php?title=%D8%AC%D8%AF%D9%8A%D8%AF%D8%A9&redirect=no
49	المحيط الأطلسي	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%85%D8%AD%D9%8A%D8%B7_%D8%A7%D9%84%D8%A3%D8%B7%D9%84%D8%B3%D9%8A
50	محافظة شمال غزة	https://ar.wikipedia.org/w/index.php?title=%D9%85%D8%AD%D8%A7%D9%81%D8%B8%D8%A9_%D8%B4%D9%85%D8%A7%D9%84_%D8%BA%D8%B2%D8%A9&redirect=no
50	مستوطنات إسرائيلية	https://ar.wikipedia.org/w/index.php?title=%D9%85%D8%B3%D8%AA%D9%88%D8%B7%D9%86%D8%A7%D8%AA_%D8%A5%D8%B3%D8%B1%D8%A7%D8%A6%D9%8A%D9%84%D9%8A%D8%A9&redirect=no
51	صاروخ بالستي	https://ar.wikipedia.org/w/index.php?title=%D8%B5%D8%A7%D8%B1%D9%88%D8%AE_%D8%A8%D8%A7%D9%84%D8%B3%D8%AA%D9%8A&redirect=no
51	أزمة كوريا الشمالية (2013)	https://ar.wikipedia.org/wiki/%D8%A3%D8%B2%D9%85%D8%A9_%D9%83%D9%88%D8%B1%D9%8A%D8%A7_%D8%A7%D9%84%D8%B4%D9%85%D8%A7%D9%84%D9%8A%D8%A9_(2013)
51	إطلاق صاروخ	https://ar.wikipedia.org/wiki/%D8%A5%D8%B7%D9%84%D8%A7%D9%82_%D8%B5%D8%A7%D8%B1%D9%88%D8%AE
52	كريستال بالاس	https://ar.wikipedia.org/wiki/%D9%83%D8%B1%D9%8A%D8%B3%D8%AA%D8%A7%D9%84_%D8%A8%D8%A7%D9%84%D8%A7%D8%B3
52	كأس إنجلترا	https://ar.wikipedia.org/w/index.php?title=%D9%83%D8%A3%D8%B3_%D8%A5%D9%86%D8%AC%D9%84%D8%AA%D8%B1%D8%A7&redirect=no
52	مانشستر يونايتد	https://ar.wikipedia.org/wiki/%D9%85%D8%A7%D9%86%D8%B4%D8%B3%D8%AA%D8%B1_%D9%8A%D9%88%D9%86%D8%A7%D9%8A%D8%AA%D8%AF
53	سوني أونلاين إنترتينمنت	https://ar.wikipedia.org/wiki/%D8%B3%D9%88%D9%86%D9%8A_%D8%A3%D9%88%D9%86%D9%84%D8%A7%D9%8A%D9%86_

Text_no	Article Title	URL
		%D8%A5%D9%86%D8%AA%D8%B1%D8%AA%D9%8A%D9%86%D9%85%D9%86%D8%AA
53	أرباح	https://ar.wikipedia.org/w/index.php?title=%D8%A3%D8%B1%D8%A8%D8%A7%D8%AD&redirect=no
53	قائمة ألعاب بلاي ستيشن	https://ar.wikipedia.org/wiki/%D9%82%D8%A7%D8%A6%D9%85%D8%A9_%D8%A3%D9%84%D8%B9%D8%A7%D8%A8_%D8%A8%D9%84%D8%A7%D9%8A_%D8%B3%D8%AA%D9%8A%D8%B4%D9%86
53	نجاح	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%AC%D8%A7%D8%AD&redirect=no
54	سامسونج	https://ar.wikipedia.org/wiki/%D8%B3%D8%A7%D9%85%D8%B3%D9%88%D9%86%D8%AC
55	الهجوم الإسرائيلي على قطاع غزة (ديسمبر 2008)	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%87%D8%AC%D9%88%D9%85_%D8%A7%D9%84%D8%A5%D8%B3%D8%B1%D8%A7%D8%A6%D9%8A%D9%84%D9%8A_%D8%B9%D9%84%D9%89_%D9%82%D8%B7%D8%A7%D8%B9_%D8%BA%D8%B2%D8%A9_(%D8%AF%D9%8A%D8%B3%D9%85%D8%A8%D8%B1_2008)&redirect=no
55	مجلس تقدير البرمجيات الترفيهية	https://ar.wikipedia.org/wiki/%D9%85%D8%AC%D9%84%D8%B3_%D8%AA%D9%82%D8%AF%D9%8A%D8%B1_%D8%A7%D9%84%D8%A8%D8%B1%D9%85%D8%AC%D9%8A%D8%A7%D8%AA_%D8%A7%D9%84%D8%AA%D8%B1%D9%81%D9%8A%D9%87%D9%8A%D8%A9
55	معلومات الألعاب الأوروبية	https://ar.wikipedia.org/wiki/%D9%85%D8%B9%D9%84%D9%88%D9%85%D8%A7%D8%AA_%D8%A7%D9%84%D8%A3%D9%84%D8%B9%D8%A7%D8%A8_%D8%A7%D9%84%D8%A3%D9%88%D8%B1%D9%88%D8%A8%D9%8A%D8%A9
56	سامسونج جالكسي	https://ar.wikipedia.org/wiki/%D8%B3%D8%A7%D9%85%D8%B3%D9%88%D9%86%D8%AC_%D8%AC%D8%A7%D9%84%D8%A7%D9%83%D8%B3%D9%8A
56	جالكسي اس المطور	https://ar.wikipedia.org/wiki/%D8%AC%D8%A7%D9%84%D9%83%D8%B3%D9%8A_%D8%A5%D8%B3_%D8%A7%D9%84%D9%85%D8%B7%D9%88%D8%B1
57	الدولار الأمريكي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AF%D9%88%D9%84%D8%A7%D8%B1_%D8%A7%D9%84%D8%A3%D9%85%D8%B1%D9%8A%D9%83%D9%8A&redirect=no
57	مدينة القاهرة	https://ar.wikipedia.org/w/index.php?title=%D9%85%D8%AF%D9%8A%D9%86%D8%A9_%D8%A7%D9%84%D9%82%D8%A7%D9%87%D8%B1%D8%A9&redirect=no
58	جماعة أبو سياف	https://ar.wikipedia.org/wiki/%D8%AC%D9%85%D8%A7%D8%B9%D8%A9_%D8%A3%D8%A8%D9%88_%D8%B3%D9%8A%D8%A7%D9%81
59	نادي الفجيرة	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D8%A7%D9%84%D9%81%D8%AC%D9%8A%D8%B1%D8%A9
59	الأهلي الإماراتي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%A3%D9%87%D9%84%D9%8A_%D8%A7%D9%84%D8%A5%D9%85%D8%A7%D8%B1%D8%A7%D8%AA%D9%8A&redirect=no
59	العين الإماراتي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%B9%D9%8A%D9%86_%D8%A7%D9%84%D8%A5%D9%85%D8%A7%D8%B1%D8%A7%D8%AA%D9%8A&redirect=no

Text_no	Article Title	URL
60	المسيحية في جنوب السودان	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%85%D8%B3%D9%8A%D8%AD%D9%8A%D8%A9_%D9%81%D9%8A_%D8%AC%D9%86%D9%88%D8%A8_%D8%A7%D9%84%D8%B3%D9%88%D8%AF%D8%A7%D9%86
61	ماتس هوملز	https://ar.wikipedia.org/wiki/%D9%85%D8%A7%D8%AA%D8%B3_%D9%87%D9%88%D9%85%D9%84%D8%B2
61	كأس السوبر الألماني	https://ar.wikipedia.org/wiki/%D9%83%D8%A3%D8%B3_%D8%A7%D9%84%D8%B3%D9%88%D8%A8%D8%B1_%D8%A7%D9%84%D8%A3%D9%84%D9%85%D8%A7%D9%86%D9%8A
61	نادي بايرن ميونيخ	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%A7%D8%AF%D9%8A_%D8%A8%D8%A7%D9%8A%D8%B1%D9%86_%D9%85%D9%8A%D9%88%D9%86%D9%8A%D8%AE&redirect=no
62	ليفربول	https://ar.wikipedia.org/wiki/%D9%84%D9%8A%D9%81%D8%B1%D8%A8%D9%88%D9%84
62	بطل كبير 6 (فيلم)	https://ar.wikipedia.org/w/index.php?title=%D8%A8%D8%B7%D9%84_%D9%83%D8%A8%D9%8A%D8%B1_6_(%D9%81%D9%8A%D9%84%D9%85)&redirect=no
63	قوقل	https://ar.wikipedia.org/w/index.php?title=%D9%82%D9%88%D9%82%D9%84&redirect=no
63	مخترع تطبيقات أندرويد	https://ar.wikipedia.org/wiki/%D9%85%D8%AE%D8%AA%D8%B1%D8%B9_%D8%AA%D8%B7%D8%A8%D9%8A%D9%82%D8%A7%D8%AA_%D8%A3%D9%86%D8%AF%D8%B1%D9%88%D9%8A%D8%AF
64	كرستيانو رونالدو	https://ar.wikipedia.org/w/index.php?title=%D9%83%D8%B1%D8%B3%D8%AA%D9%8A%D8%A7%D9%86%D9%88_%D8%B1%D9%88%D9%86%D8%A7%D9%84%D8%AF%D9%88&redirect=no
64	زين الدين زيدان	https://ar.wikipedia.org/wiki/%D8%B2%D9%8A%D9%86_%D8%A7%D9%84%D8%AF%D9%8A%D9%86_%D8%B2%D9%8A%D8%AF%D8%A7%D9%86
65	أنغيلا ميركل	https://ar.wikipedia.org/wiki/%D8%A3%D9%86%D8%BA%D9%8A%D9%84%D8%A7_%D9%85%D9%8A%D8%B1%D9%83%D9%84
65	الناتو (حلف شمال الأطلنطي)	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%86%D8%A7%D8%AA%D9%88_(%D8%AD%D9%84%D9%81_%D8%B4%D9%85%D8%A7%D9%84_%D8%A7%D9%84%D8%A3%D8%B7%D9%84%D9%86%D8%B7%D9%8A)&redirect=no
66	دمشق (محافظة)	https://ar.wikipedia.org/w/index.php?title=%D8%AF%D9%85%D8%B4%D9%82_(%D9%85%D8%AD%D8%A7%D9%81%D8%B8%D8%A9)&redirect=no
66	أيوبيو حلب	https://ar.wikipedia.org/wiki/%D8%A3%D9%8A%D9%88%D8%A8%D9%8A%D9%88_%D8%AD%D9%84%D8%A8
66	حصار اللاذقية (2011)	https://ar.wikipedia.org/w/index.php?title=%D8%AD%D8%B5%D8%A7%D8%B1_%D8%A7%D9%84%D9%84%D8%A7%D8%B0%D9%82%D9%8A%D8%A9_(2011)&redirect=no
67	معاداة السامية	https://ar.wikipedia.org/wiki/%D9%85%D8%B9%D8%A7%D8%AF%D8%A7%D8%A9_%D8%A7%D9%84%D8%B3%D8%A7%D9%85%D9%8A%D8%A9
67	الإمبراطورية بريطانية	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%A5%D9%85%D8%A8%D8%B1%D8%A7%D8%B7%D9%88%D8%B1%D9%8A%D8%A9_%D8%A8%D8%B1%D9%8A%D8%B7%D8%A7%D9%86%D9%8A%D8%A9&redirect=no

Text_no	Article Title	URL
68	هاشتاغ	https://ar.wikipedia.org/w/index.php?title=%D9%87%D8%A7%D8%B4%D8%AA%D8%A7%D8%BA&redirect=no
68	تغريدة	https://ar.wikipedia.org/w/index.php?title=%D8%AA%D8%BA%D8%B1%D9%8A%D8%AF%D8%A9&redirect=no
68	تويتر	https://ar.wikipedia.org/wiki/%D8%AA%D9%88%D9%8A%D8%AA%D8%B1
68	معركة حلب (2012–الآن)	https://ar.wikipedia.org/wiki/%D9%85%D8%B9%D8%B1%D9%83%D8%A9_%D8%AD%D9%84%D8%A8_(2012%E2%80%93%D8%A7%D9%84%D8%A2%D9%86)
69	ألساندرو دل بييرو	https://ar.wikipedia.org/w/index.php?title=%D8%A3%D9%84%D8%B3%D8%A7%D9%86%D8%AF%D8%B1%D9%88_%D8%AF%D9%84_%D8%A8%D9%8A%D9%8A%D8%B1%D9%88&redirect=no
69	ليستر سيتي	https://ar.wikipedia.org/wiki/%D9%84%D9%8A%D8%B3%D8%AA%D8%B1_%D8%B3%D9%8A%D8%AA%D9%8A
71	الجمعة العظيمة	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AC%D9%85%D8%B9%D8%A9_%D8%A7%D9%84%D8%B9%D8%B8%D9%8A%D9%85%D8%A9&redirect=no
71	المسيحيون في الناصرة	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%85%D8%B3%D9%8A%D8%AD%D9%8A%D9%88%D9%86_%D9%81%D9%8A_%D8%A7%D9%84%D9%86%D8%A7%D8%B5%D8%B1%D8%A9
71	القدس في المسيحية	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%82%D8%AF%D8%B3_%D9%81%D9%8A_%D8%A7%D9%84%D9%85%D8%B3%D9%8A%D8%AD%D9%8A%D8%A9
72	ساعة بيج بن	https://ar.wikipedia.org/w/index.php?title=%D8%B3%D8%A7%D8%B9%D8%A9_%D8%A8%D9%8A%D8%AC_%D8%A8%D9%86&redirect=no
72	بيج بن عدن	https://ar.wikipedia.org/wiki/%D8%A8%D8%AC_%D8%A8%D9%86_%D8%B9%D8%AF%D9%86
73	الاحتلال الإسرائيلي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%A7%D8%AD%D8%AA%D9%84%D8%A7%D9%84_%D8%A7%D9%84%D8%A5%D8%B3%D8%B1%D8%A7%D8%A6%D9%8A%D9%84%D9%8A&redirect=no
73	جبل المكبر	https://ar.wikipedia.org/wiki/%D8%AC%D8%A8%D9%84_%D8%A7%D9%84%D9%85%D9%83%D8%A8%D8%B1
74	مقاتلة الجيل الخامس	https://ar.wikipedia.org/w/index.php?title=%D9%85%D9%82%D8%A7%D8%AA%D9%84%D8%A9_%D8%A7%D9%84%D8%AC%D9%8A%D9%84_%D8%A7%D9%84%D8%AE%D8%A7%D9%85%D8%B3&redirect=no
74	بحر البلطيق	https://ar.wikipedia.org/wiki/%D8%A8%D8%AD%D8%B1_%D8%A7%D9%84%D8%A8%D9%84%D8%B7%D9%8A%D9%82
74	مدن إتحادية روسية	https://ar.wikipedia.org/w/index.php?title=%D9%85%D8%AF%D9%86_%D8%A5%D8%AA%D8%AD%D8%A7%D8%AF%D9%8A%D8%A9_%D8%B1%D9%88%D8%B3%D9%8A%D8%A9&redirect=no
75	مستقبل مانوز	https://ar.wikipedia.org/wiki/%D9%85%D8%B3%D8%AA%D9%82%D8%A8%D9%84_%D9%85%D8%A7%D9%86%D9%88%D8%B2
75	مستقبل الروبوتات	https://ar.wikipedia.org/w/index.php?title=%D9%85%D8%B3%D8%AA%D9%82%D8%A8%D9%84_%D8%A7%D9%84%D8%B1%D9%88%D8%A8%D9%88%D8%AA%D8%A7%D8%AA&redirect=no

Text_no	Article Title	URL
75	علم الجينوم البنيوي	https://ar.wikipedia.org/wiki/%D8%B9%D9%84%D9%85_%D8%A7%D9%84%D8%AC%D9%8A%D9%86%D9%88%D9%85_%D8%A7%D9%84%D8%A8%D9%86%D9%8A%D9%88%D9%8A
75	قائمة تطبيقات تقنية النانو	https://ar.wikipedia.org/w/index.php?title=%D9%82%D8%A7%D8%A6%D9%85%D8%A9_%D8%AA%D8%B7%D8%A8%D9%8A%D9%82%D8%A7%D8%AA_%D8%AA%D9%82%D9%86%D9%8A%D8%A9_%D8%A7%D9%84%D9%86%D8%A7%D9%86%D9%88&redirect=no
76	حصار مدينة الصدر	https://ar.wikipedia.org/wiki/%D8%AD%D8%B5%D8%A7%D8%B1_%D9%85%D8%AF%D9%8A%D9%86%D8%A9_%D8%A7%D9%84%D8%B5%D8%AF%D8%B1
76	المنطقة الخضراء	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D9%85%D9%86%D8%B7%D9%82%D8%A9_%D8%A7%D9%84%D8%AE%D8%B6%D8%B1%D8%A7%D8%A1
77	مروحية متعددة الأغراض	https://ar.wikipedia.org/wiki/%D9%85%D8%B1%D9%88%D8%AD%D9%8A%D8%A9_%D9%85%D8%AA%D8%B9%D8%AF%D8%AF%D8%A9_%D8%A7%D9%84%D8%A3%D8%BA%D8%B1%D8%A7%D8%B6
78	واقع معزز	https://ar.wikipedia.org/wiki/%D9%88%D8%A7%D9%82%D8%B9_%D9%85%D8%B9%D8%B2%D8%B2
78	نقل البيانات	https://ar.wikipedia.org/wiki/%D9%86%D9%82%D9%84_%D8%A7%D9%84%D8%A8%D9%8A%D8%A7%D9%86%D8%A7%D8%AA
78	نقل الجينات الأفقي	https://ar.wikipedia.org/w/index.php?title=%D9%86%D9%82%D9%84_%D8%A7%D9%84%D8%AC%D9%8A%D9%86%D8%A7%D8%AA_%D8%A7%D9%84%D8%A3%D9%81%D9%82%D9%8A&redirect=no
78	مرسال الحمض النووي الريبي	https://ar.wikipedia.org/w/index.php?title=%D9%85%D8%B1%D8%B3%D8%A7%D9%84_%D8%A7%D9%84%D8%AD%D9%85%D8%B6_%D8%A7%D9%84%D9%86%D9%88%D9%88%D9%8A_%D8%A7%D9%84%D8%B1%D9%8A%D8%A8%D9%8A&redirect=no
79	معركة تلعفر 2014	https://ar.wikipedia.org/wiki/%D9%85%D8%B9%D8%B1%D9%83%D8%A9_%D8%AA%D9%84%D8%B9%D9%81%D8%B1_2014
79	البشمركة	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%A8%D8%B4%D9%85%D8%B1%D9%83%D8%A9&redirect=no
80	الجامعة الأميركية	https://ar.wikipedia.org/wiki/%D8%A7%D9%84%D8%AC%D8%A7%D9%85%D8%B9%D8%A9_%D8%A7%D9%84%D8%A3%D9%85%D9%8A%D8%B1%D9%83%D9%8A%D8%A9
80	باراك أوباما	https://ar.wikipedia.org/wiki/%D8%A8%D8%A7%D8%B1%D8%A7%D9%83_%D8%A3%D9%88%D8%A8%D8%A7%D9%85%D8%A7
80	نيوزويك	https://ar.wikipedia.org/wiki/%D9%86%D9%8A%D9%88%D8%B2%D9%88%D9%8A%D9%83
80	مقال	https://ar.wikipedia.org/wiki/%D9%85%D9%82%D8%A7%D9%84
80	بشار الأسد	https://ar.wikipedia.org/wiki/%D8%A8%D8%B4%D8%A7%D8%B1_%D8%A7%D9%84%D8%A3%D8%B3%D8%AF
80	الرئيس الأمريكي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%B1%D8%A6%D9%8A%D8%B3_%D8%A7%D9%84%D8%A3%D9%85%D8%B1%D9%8A%D9%83%D9%8A&redirect=no

Text_no	Article Title	URL
81	مدينة كيبيك	https://ar.wikipedia.org/wiki/%D9%85%D8%AF%D9%8A%D9%86%D8%A9_%D9%83%D9%8A%D8%A8%D9%83
81	فلاديمير بوتين	https://ar.wikipedia.org/wiki/%D9%81%D9%84%D8%A7%D8%AF%D9%8A%D9%85%D9%8A%D8%B1_%D8%A8%D9%88%D8%AA%D9%8A%D9%86
81	ماريك سوتشي	https://ar.wikipedia.org/wiki/%D9%85%D8%A7%D8%B1%D9%8A%D9%83_%D8%B3%D9%88%D8%AA%D8%B4%D9%8A
81	كأس الكرملين	https://ar.wikipedia.org/wiki/%D9%83%D8%A3%D8%B3_%D8%A7%D9%84%D9%83%D8%B1%D9%85%D9%84%D9%8A%D9%86
82	شيعة العراق	https://ar.wikipedia.org/wiki/%D8%B4%D9%8A%D8%B9%D8%A9_%D8%A7%D9%84%D8%B9%D8%B1%D8%A7%D9%82
84	بشار الأسد	https://ar.wikipedia.org/wiki/%D8%A8%D8%B4%D8%A7%D8%B1_%D8%A7%D9%84%D8%A3%D8%B3%D8%AF
84	شعارات مقاطعات هولندا	https://ar.wikipedia.org/wiki/%D8%B4%D8%B9%D8%A7%D8%B1%D8%A7%D8%AA_%D9%85%D9%82%D8%A7%D8%B7%D8%B9%D8%A7%D8%AA_%D9%87%D9%88%D9%84%D9%86%D8%AF%D8%A7
84	مؤشر بورصة أمستردام	https://ar.wikipedia.org/wiki/%D9%85%D8%A4%D8%B4%D8%B1_%D8%A8%D9%88%D8%B1%D8%B5%D8%A9_%D8%A3%D9%85%D8%B3%D8%AA%D8%B1%D8%AF%D8%A7%D9%85
85	نادي بوروسيا مونشنغلادباخ	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%A7%D8%AF%D9%8A_%D8%A8%D9%88%D8%B1%D9%88%D8%B3%D9%8A%D8%A7_%D9%85%D9%88%D9%86%D8%B4%D9%86%D8%BA%D9%84%D8%A7%D8%AF%D8%A8%D8%A7%D8%AE&redirect=no
85	الدوري الألماني الدرجة الأولى	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AF%D9%88%D8%B1%D9%8A_%D8%A7%D9%84%D8%A3%D9%84%D9%85%D8%A7%D9%86%D9%8A_%D8%A7%D9%84%D8%AF%D8%B1%D8%AC%D8%A9_%D8%A7%D9%84%D8%A3%D9%88%D9%84%D9%89&redirect=no
86	ستوك سيتي	https://ar.wikipedia.org/wiki/%D8%B3%D8%AA%D9%88%D9%83_%D8%B3%D9%8A%D8%AA%D9%8A
86	نادي سندرلاند	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D8%B3%D9%86%D8%AF%D8%B1%D9%84%D8%A7%D9%86%D8%AF
87	غاريث بيل	https://ar.wikipedia.org/wiki/%D8%BA%D8%A7%D8%B1%D9%8A%D8%AB_%D8%A8%D9%8A%D9%84
87	ريال سوسييداد	https://ar.wikipedia.org/w/index.php?title=%D8%B1%D9%8A%D8%A7%D9%84_%D8%B3%D9%88%D8%B3%D9%8A%D9%8A%D8%AF%D8%A7%D8%AF&redirect=no
87	الليغا	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%84%D9%8A%D8%BA%D8%A7&redirect=no
87	نادي ريال مدريد	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%A7%D8%AF%D9%8A_%D8%B1%D9%8A%D8%A7%D9%84_%D9%85%D8%AF%D8%B1%D9%8A%D8%AF&redirect=no
88	الدوري الإنجليزي الممتاز 14-2013	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AF%D9%88%D8%B1%D9%8A_%D8%A7%D9%84%D8%A5%D9%86%D8%AC%D9%84%D9%8A%D8%B2%D9%8A_%D8%A7%D9%84%D9%85%D9%85%D8%AA%D8%A7%D8%B2_2013-14&redirect=no

Text_no	Article Title	URL
88	نادي أستون فيلا	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%A7%D8%AF%D9%8A_%D8%A3%D8%B3%D8%AA%D9%88%D9%86_%D9%81%D9%8A%D9%84%D8%A7&redirect=no
89	نظام حجز إلكتروني	https://ar.wikipedia.org/wiki/%D9%86%D8%B8%D8%A7%D9%85_%D8%AD%D8%AC%D8%B2_%D8%A5%D9%84%D9%83%D8%AA%D8%B1%D9%88%D9%86%D9%8A
89	شركة الإتصالات السعودية	https://ar.wikipedia.org/w/index.php?title=%D8%B4%D8%B1%D9%83%D8%A9_%D8%A7%D9%84%D8%A5%D8%AA%D8%B5%D8%A7%D9%84%D8%A7%D8%AA_%D8%A7%D9%84%D8%B3%D8%B9%D9%88%D8%AF%D9%8A%D8%A9&redirect=no
90	الدوري الفرنسي	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AF%D9%88%D8%B1%D9%8A_%D8%A7%D9%84%D9%81%D8%B1%D9%86%D8%B3%D9%8A&redirect=no
91	فؤاد معصوم	https://ar.wikipedia.org/wiki/%D9%81%D8%A4%D8%A7%D8%AF_%D9%85%D8%B9%D8%B5%D9%88%D9%85
92	يوروفيجن	https://ar.wikipedia.org/w/index.php?title=%D9%8A%D9%88%D8%B1%D9%88%D9%81%D9%8A%D8%AC%D9%86&redirect=no
92	علم فلسطين	https://ar.wikipedia.org/wiki/%D8%B9%D9%84%D9%85_%D9%81%D9%84%D8%B3%D8%B7%D9%8A%D9%86
93	إسطنبول (محافظة)	https://ar.wikipedia.org/wiki/%D8%A5%D8%B3%D8%B7%D9%86%D8%A8%D9%88%D9%84_(%D9%85%D8%AD%D8%A7%D9%81%D8%B8%D8%A9)
93	عيد العمال العالمي	https://ar.wikipedia.org/w/index.php?title=%D8%B9%D9%8A%D8%AF_%D8%A7%D9%84%D8%B9%D9%85%D8%A7%D9%84_%D8%A7%D9%84%D8%B9%D8%A7%D9%84%D9%85%D9%8A&redirect=no
94	مياه شرب	https://ar.wikipedia.org/w/index.php?title=%D9%85%D9%8A%D8%A7%D9%87_%D8%B4%D8%B1%D8%A8&redirect=no
94	خصائص الماء	https://ar.wikipedia.org/wiki/%D8%AE%D8%B5%D8%A7%D8%A6%D8%B5_%D8%A7%D9%84%D9%85%D8%A7%D8%A1
95	عيد الفصح المسيحي	https://ar.wikipedia.org/w/index.php?title=%D8%B9%D9%8A%D8%AF_%D8%A7%D9%84%D9%81%D8%B5%D8%AD_%D8%A7%D9%84%D9%85%D8%B3%D9%8A%D8%AD%D9%8A&redirect=no
95	قداس	https://ar.wikipedia.org/w/index.php?title=%D9%82%D8%AF%D8%A7%D8%B3&redirect=no
96	خليل الوزير	https://ar.wikipedia.org/wiki/%D8%AE%D9%84%D9%8A%D9%84_%D8%A7%D9%84%D9%88%D8%B2%D9%8A%D8%B1
96	عبد العزيز بوتفليقة	https://ar.wikipedia.org/wiki/%D8%B9%D8%A8%D8%AF_%D8%A7%D9%84%D8%B9%D8%B2%D9%8A%D8%B2_%D8%A8%D9%88%D8%AA%D9%81%D9%84%D9%8A%D9%82%D8%A9
96	شكيب خليل	https://ar.wikipedia.org/wiki/%D8%B4%D9%83%D9%8A%D8%A8_%D8%AE%D9%84%D9%8A%D9%84
97	نادي مانشستر سيتي	https://ar.wikipedia.org/w/index.php?title=%D9%86%D8%A7%D8%AF%D9%8A_%D9%85%D8%A7%D9%86%D8%B4%D8%B3%D8%AA%D8%B1_%D8%B3%D9%8A%D8%AA%D9%8A&redirect=no
97	آرسنال	https://ar.wikipedia.org/w/index.php?title=%D8%A2%D8%B1%D8%B3%D9%86%D8%A7%D9%84&redirect=no
98	مارين لوبان	https://ar.wikipedia.org/wiki/%D9%85%D8%A7%D8%B1%D9%8A%D9%86_%D9%84%D9%88%D8%A8%D8%A7%D9%86

Text_no	Article Title	URL
99	سناب شات	https://ar.wikipedia.org/wiki/%D8%B3%D9%86%D8%A7%D8%A8_%D8%B4%D8%A7%D8%AA
99	المملكة العربية السعودية	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D9%85%D9%85%D9%84%D9%83%D8%A9_%D8%A7%D9%84%D8%B9%D8%B1%D8%A8%D9%8A%D8%A9_%D8%A7%D9%84%D8%B3%D8%B9%D9%88%D8%AF%D9%8A%D8%A9&redirect=no
100	نادي لاتسيو	https://ar.wikipedia.org/wiki/%D9%86%D8%A7%D8%AF%D9%8A_%D9%84%D8%A7%D8%AA%D8%B3%D9%8A%D9%88
100	الدوري الإيطالي الممتاز	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%84%D8%AF%D9%88%D8%B1%D9%8A_%D8%A7%D9%84%D8%A5%D9%8A%D8%B7%D8%A7%D9%84%D9%8A_%D8%A7%D9%84%D9%85%D9%85%D8%AA%D8%A7%D8%B2&redirect=no
100	انتر ميلان	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D9%86%D8%AA%D8%B1_%D9%85%D9%8A%D9%84%D8%A7%D9%86&redirect=no

Appendix C: Precision, and Recall Calculations

To illustrate how we compute the precision and recall for any input text, we give the following example: assume that the end user is searching for the short text "Hundreds demonstrate in solidarity with Aleppo in Istanbul" "المئات يتظاهرون تضامناً مع حلب في اسطنبول".

The following table shows the expected relevant articles for the input text:

Table C.1: Relevant articles for input text

Relevant articles for text: "المئات يتظاهرون تضامناً مع حلب في اسطنبول"	
اسطنبول	https://ar.wikipedia.org/w/index.php?title=%D8%A7%D8%B3%D8%B7%D9%86%D8%A8%D9%88%D9%84&redirect=no
حلب	https://ar.wikipedia.org/wiki/%D8%AD%D9%84%D8%A8

While the system returns the article "اسطنبول", based on the relevant articles and the returned articles we calculated the precision and recall for each input text as following:

$$\textit{(1 + 0) = 1}\boldsymbol{\textit{Precision}} = \frac{\textit{Total number of terms linked to relevant Wikipedia articles}}{\textit{Total number of terms linked to Wikipedia articles}} = 1/$$

$$\boldsymbol{\textit{Recall}} = \frac{\textit{Total number of terms linked to relevant Wikipedia articles}}{\textit{Total number of terms that can be linked to Wikipedia articles}} = 1/(1 + 1)$$

$$= 0.5$$

The following table shows precision, recall and F-measure for each input text of the dataset.

Text_No	Precision Calculations	Precision	Recall Calculations	Recall	F-measure
1	2/(2+0)	1	2/(2+0)	1	1
2	2/(2+3)	0.4	2/(2+2)	0.5	0.44
3	2/(2+0)	1	2/(2+0)	1	1
4	2/(2+1)	0.67	2/(2+0)	1	0.8
5	4/(4+1)	0.8	4/(4+0)	1	0.89
6	1/(1+0)	1	1/(1+0)	1	1
7	3/(3+1)	0.75	3/(3+1)	0.75	0.75
8	0/(0+1)	0	0/(0+1)	0	0
9	1/(1+0)	1	1/(1+1)	0.5	0.67
10	1/(1+0)	1	1/(1+0)	1	1
11	1/(1+0)	1	1/(1+2)	0.33	0.5
12	2/(2+0)	1	2/(2+0)	1	1
13	1/(1+2)	0.333333	1/(1+1)	0.5	0.4
14	1/(1+0)	1	1/(1+0)	1	1
15	2/(2+2)	0.5	2/(2+0)	1	0.67
16	1/(1+0)	1	1/(1+0)	1	1
17	0/(0+0)	0	0/(0+2)	0	0
18	3/(3+0)	1	3/(3+0)	1	1
19	2/(2+0)	1	2/(2+0)	1	1
20	3/(3+4)	0.43	3/(3+0)	1	0.6
21	1/(1+0)	1	1/(1+2)	0.33	0.5
22	1/(1+0)	1	1/(1+0)	1	1
23	3/(3+0)	1	3/(3+0)	1	1
24	2/(2+0)	1	2/(2+1)	0.67	0.8
25	3/(3+0)	1	3/(3+0)	1	1
26	1/(1+0)	1	1/(1+1)	0.5	0.67
27	1/(1+1)	0.5	2/(2+0)	1	0.67
28	2/(2+1)	0.67	2/(2+0)	1	0.8
29	1/(1+0)	1	1/(1+0)	1	1
30	2/(2+0)	1	2/(2+0)	1	1
31	2/(2+0)	1	2/(2+0)	1	1
32	1/(1+0)	1	1/(1+0)	1	1
33	1/(1+0)	1	1/(1+0)	1	1
34	2/(2+0)	1	2/(2+0)	1	1
35	1/(1+1)	0.5	1/(1+0)	1	0.67
36	2/(2+0)	1	2/(2+0)	1	1
37	2/(2+1)	0.67	2/(2+0)	1	0.8
38	0/(0+2)	0	0/(0+0)	0	0
39	2/(2+0)	1	2/(2+0)	1	1
40	2/(2+3)	0.4	2/(2+0)	1	0.57
41	3/(3+1)	0.75	3/(3+0)	1	0.86
42	2/(2+2)	0.5	2/(2+0)	1	0.67

Text_No	Precision Calculations	Precision	Recall Calculations	Recall	F-measure
43	1/(1+1)	0.5	1/(1+1)	0.5	0.5
44	1/(1+2)	0.33	1/(1+1)	0.5	0.4
45	2/(2+0)	1	2/(2+0)	1	1
46	2/(2+1)	0.67	3/(3+0)	1	0.8
47	2/(2+0)	1	2/(2+0)	1	1
48	2/(2+0)	1	2/(2+0)	1	1
49	3/(3+3)	0.5	3/(3+0)	1	0.67
50	1/(1+1)	0.5	1/(1+0)	1	0.67
51	2/(2+1)	0.67	2/(2+0)	1	0.8
52	3/(3+0)	1	3/(3+1)	0.75	0.86
53	2/(2+2)	0.5	2/(2+0)	1	0.67
54	1/(1+0)	1	1/(1+0)	1	1
55	0/(0+3)	0	0/(0+0)	0	0
56	1/(1+1)	0.5	1/(1+0)	1	0.67
57	2/(2+0)	1	2/(2+0)	1	1
58	1/(1+0)	1	1/(1+0)	1	1
59	3/(3+0)	1	3/(3+0)	1	1
60	1/(1+0)	1	1/(1+0)	1	1
61	2/(2+1)	0.67	2/(2+0)	1	0.8
62	1/(1+1)	0.5	1/(1+0)	1	0.67
63	2/(2+0)	1	2/(2+0)	1	1
64	2/(2+0)	1	2/(2+0)	1	1
65	2/(2+0)	1	2/(2+1)	0.666667	0.8
66	3/(3+0)	1	3/(3+0)	1	1
67	1/(1+1)	0.5	1/(1+0)	1	0.67
68	4/(4+0)	1	4/(4+1)	0.8	0.89
69	2/(2+0)	1	2/(2+0)	1	1
70	0/(0+0)	0	0/(0+1)	0	0
71	2/(2+1)	0.67	2/(2+0)	1	0.8
72	1/(1+1)	0.5	1/(1+0)	1	0.67
73	2/(2+0)	1	2/(2+1)	0.666667	0.8
74	1/(1+2)	0.333333	1/(1+0)	1	0.5
75	3/(3+1)	0.75	3/(3+0)	1	0.86
76	2/(2+0)	1	2/(2+1)	0.666667	0.8
77	0/(0+0)	0	0/(0+0)	0	0
78	3/(3+1)	0.75	3/(3+0)	1	0.86
79	1/(1+1)	0.5	1/(1+1)	0.5	0.5
80	3/(3+3)	0.5	3/(3+0)	1	0.67
81	2/(2+2)	0.5	2/(2+1)	0.67	0.57
82	1/(1+0)	1	1/(1+1)	0.5	0.67
83	0/(0+0)	0	0/(0+2)	0	0
84	1/(1+2)	0.33	1/(1+1)	0.5	0.4

Text_No	Precision Calculations	Precision	Recall Calculations	Recall	F-measure
85	2/(2+0)	1	2/(2+1)	0.67	0.8
86	2/(2+0)	1	2/(2+0)	1	1
87	4/(4+0)	1	4/(4+0)	1	1
88	2/(2+0)	1	2/(2+1)	0.67	0.8
89	1/(1+1)	0.5	1/(1+0)	1	0.67
90	1/(1+0)	1	1/(1+1)	0.5	0.67
91	1/(1+0)	1	1/(1+1)	0.5	0.67
92	2/(2+0)	1	2/(2+0)	1	1
93	2/(2+0)	1	2/(2+0)	1	1
94	1/(1+1)	0.5	1/(1+0)	1	0.67
95	1/(1+1)	0.5	1/(1+1)	0.5	0.5
96	2/(2+1)	0.67	2/(2+0)	1	0.8
97	2/(2+0)	1	2/(2+0)	1	1
98	0/(0+1)	0	0/(0+0)	0	0
99	2/(2+0)	1	2/(2+0)	1	1
100	3/(3+0)	1	3/(3+0)	1	1