UNIVERSITI PUTRA MALAYSIA

ADAPTIVE MODEL FOR SEMANTIC QUESTION ANSWERING DISAMBIGUATION OVER LINKED DATA

HAZRINA BINTI SOFIAN

FSKTM 2018 66

ADAPTIVE MODEL FOR SEMANTIC QUESTION ANSWERING DISAMBIGUATION OVER LINKED DATA

By

HAZRINA BINTI SOFIAN

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfillment of the Requirements for the Degree of Doctor of Philosophy

April 2018

DEDICATION

Dedicated to human beings

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

ADAPTIVE MODEL FOR SEMANTIC QUESTION ANSWERING DISAMBIGUATION OVER LINKED DATA

By

HAZRINA BINTI SOFIAN

April 2018

Chairman : Associate Professor Nurfadhlina Binti Mohd Sharef, PhD
Faculty : Computer Science and Information Technology

Semantic Question Answering (SQA) accepts natural language question (NL) from users and presents the exact answer retrieved from the linked data. It requires three disambiguations which are NL question disambiguation, linked data environment disambiguation and multi-types of word disambiguation. Firstly, the NL disambiguation involves the disambiguation of three meta-mapping aspects which are the variation of question pattern, question complexity and linguistic terminologies of NL questions posed by users. Secondly, the linked data disambiguation involves the disambiguation of another four meta-mapping aspects which are the variation of datatype, resource heterogeneity, knowledge-based (KB) concept terminology and the variation of structure in the linked data. Thirdly, the word disambiguation involves the disambiguation between the linguistic terminology and the KB concept terminology. These three disambiguations are needed to be addressed simultaneously because through empirical study that had been carried out, this research has found that the Simple Protocol and RDF Query Language (SPARQL) components are determined by these seven meta-mapping aspects.

Most existing researches modify the question, manually; select only certain patterns of NL questions or select only simple questions from the dataset. Moreover, certain processes are semi-automated as some SQAs rely heavily on pre-determined lexicon knowledge for word disambiguation or manually annotate mapping for the SPARQL query constructions. However, the manual or semi-automated process is unable to cater for new question patterns posed by users or to adapt the contents in the linked data that is ever-changing and incrementally growing.

These motivate this research to firstly design the Adaptive-based Natural Language Disambiguation (ANLD) model which is integrated with the Linguistic-based SPARQL Translation Model (LBSTM), selective (Part of Speech Tagging) POS tag extraction technique, composition of syntactic representation technique and model matching technique to disambiguate NL questions. Next, this research designs the Adaptive-based Linked Data Structure Disambiguation (ALID) model that is executed if the output of the ANLD model is not able to retrieve answer from the linked data. ALID uses component-based approach and feedback loop approach to disambiguate linked data environment and to disambiguate the word ambiguity.

Precision, recall and f-measure are used as performance metrics to evaluate the accuracy of the SPARQL queries which are the outputs of this research. The accuracy is evaluated by comparing the constructed SPARQL queries with the golden standard results provided by the dataset. These results illustrate that the adaptive models are able to perform the three SQA disambiguation abilities simultaneously without manual modification. These achievements empower autonomous processing of translating NL questions to the SPARQL queries that involves users with unpredictable style of question writings against the linked data that is incrementally growing in terms of size and complexity.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

MODEL ADAPTIF UNTUK DISAMBIGUASI DALAM SISTEM SOAL JAWAB SEMANTIK KE ATAS LINKED DATA

Oleh

HAZRINA BINTI SOFIAN

April 2018

Pengerusi : Profesor Madya Nurfadhlina Binti Mohd Sharef, PhD
Fakulti : Sains Komputer dan Teknologi Maklumat

‘Semantic question answering’ (SQA) menerima soalan dalam bentuk ayat daripada pengguna dan mempamerkan jawapan yang diambil daripada linked data. SQA memerlukan tiga kemahiran disambiguasi iaitu disambiguasi soalan pengguna, disambiguasi persekitaran ‘linked data’, dan disambiguasi terminologi. Pertama, disambiguasi soalan pengguna meliputi tiga aspek meta-mapping iaitu kepelbagaian struktur ayat yang dibina oleh pengguna, kompleksiti soalan dan pemilihan terminologi oleh pengguna di dalam soalan. Kedua, disambiguasi persekitaran ‘linked data’ yang meliputi empat aspek meta-mapping iaitu kepelbagaian ‘datatype’, kepelbagaian sumber maklumat, terminologi yang disimpan di dalam konsep ‘knowledge-based’ (KB) dan kepelbagaian struktur susunan jawapan yang terdapat di dalam ‘linked data’. Ketiga, disambiguasi di antara terminologi pengguna dan terminologi di dalam KB. Ketiga-tiga disambiguasi ini perlu selesaikan secara serentak kerana menerusi kajian empirikal yang telah dijalankan, penyelidikan ini mendapati komponen Simple Protocol and RDF Query Language (SPARQL) ditentukan melalui tujuh aspek meta-mapping ini.

Kebanyakan hasil penyelidikan yang sedia ada menyelesaikan masalah ini dengan mengubah-suai struktur ayat yang dibina oleh pengguna secara manual: dengan hanya memilih struktur ayat yang tertentu atau dengan hanya memilih struktur ayat yang ringkas sahaja. Selain itu, proses-proses tertentu adalah berbentuk separa automatik kerana sesetengah SQA terlalu bergantung pada pengetahuan leksikon yang telah ditentukan untuk disambiguasi perkataan atau pemetaan ‘annotate’ secara manual untuk pembinaan pertanyaan bercirikan SPARQL. Walau bagaimanapun, proses manual atau separa-automatik ini tidak dapat memenuhi pola pertanyaan baru yang ditimbulkan oleh pengguna atau tidak dapat mengadaptasi kandungan dalam ‘linked data’ yang sentiasa berubah dan sentiasa bertambah secara berperingkat.

Perkara-perkara ini memberi motivasi kepada penyelidikan ini untuk mereka-bentuk 'Adaptive-based Natural Language Disambiguation (ANLD) model' yang diintegrasikan dengan 'Linguistic-based SPARQL Translation Model' (LBSTM), 'selective (Part of Speech Tagging) POS tag extraction technique', 'composition of syntactic representation technique' dan 'model matching technique' untuk mengdisambiguasikan soalan pengguna. Seterusnya, penyelidikan ini mereka-bentuk 'Adaptive-based Linked Data Structure Disambiguation (ALID) model' yang diaktifkan sekiranya output kepada ANLD model tidak berjaya mencapai data dari 'linked data'. ALID menggunakan 'component-based approach' dan 'feedback loop approach' untuk mengdisambiguasikan persekitaran 'linked data' dan disambiguasi terminologi.

'Precision', 'recall' dan 'f-measure' adalah penanda arasan prestasi yang digunakan untuk menilai ketepatan output penyelidikan ini iaitu SPARQL query. Ketepatan output ini boleh dinilai berdasarkan perbandingan SPARQL query yang terbina dengan keputusan 'golden standard' yang dibekalkan oleh 'dataset'. Keputusan tentang ketepatan output ini menunjukkan bahawa solusi kajian ini mampu melaksanakan ketiga-tiga kemahiran disambiguasi SQA secara serentak dan tanpa memerlukan pengubahsuaian secara manual. Pencapaian ini memberi kebolehan pemprosesan autonomi untuk menterjemahkan soalan bahasa semulajadi kepada SPARQL query yang melibatkan pengguna menggunakan gaya penulisan soalan yang tidak diketahui. Persekitaran meliputi struktur pembinaan ayat yang tidak dapat dijangkakan dan persekitaran 'linked data' yang semakin berkembang dari segi saiz dan kompleksiti.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Allah S.W.T for giving me the strength, health and courage to complete this research.

I would also like to convey my deepest appreciation to Associate Professor Dr. Nurfadhlina Binti Mohd Sharef and my co-supervisors Professor Hamidah Binti Ibrahim and Associate Professor Dr. Masrah Azrifah Binti Azmi Murad for their time, motivations, close supervisions, encouragements and stimulating suggestions throughout the duration of this study. I have furthermore to thank the Faculty of Computer Science and Information Technology for giving me the permission and the opportunity to commence this research.

My sincere thanks also go to Mybrain15, Malaysian Higher Education and the Fundamental Research Grant Scheme (FRGS) (project number: FRGS/2/2013/ICT07/UPM/02/1) Ministry of Higher Education, Malaysia for providing the scholarship and funding for this research. Special thanks to the members of FRGS committee: Professor Dr. Shahrul Azman Bin Mohd Noah, Associate Professor Datin Dr. Norwati Binti Mustapha and Dr. Rabiah Binti Abdul Kadir.

To all the reviewers of the Component-based Evaluation Mechanism; thank you for your efforts and interests to understand and provide me with valuable feedbacks.

To all of my friends in UPM especially Arzilawati, Yana, Harnani, Ana Salwa, Raihani, Saiful Nizam, Azlan Shah, Hisham, Rafiez, Aliyu, Khaleed, Esmail and Muath; thank you for being great companions whilst this research is being carried out. To my best friend, Sarah and my beloved aunty Rosni Binti Aub, thank you for your endless encouragements.

I would also like to express my deepest gratitude to my parents; Sharifah Robiah Binti Syed Yusoff and Sofian Bin Shafie who never failed to pray for my success. Your words of encouragements had given me the confidence to carry out and finish this entire research within the stipulated time.

As my utmost gratification, I would like to give my special thanks to my beloved husband, Mohd Noorshah Bin Mohd Izhar whose patient and love had taught me the true meaning of perseverance throughout this journey. Last but not least, thank you to both my sons: Muhammad Adam Hadif and Muhammad Hariz Hadif for allowing me to take some of my time away from them to complete this research.

I certify that a Thesis Examination Committee has met on 12 April 2018 to conduct the final examination of Hazrina binti Sofian on her thesis entitled "Adaptive Model for Semantic Question Answering Disambiguation Over Linked Data" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Nor Fazlida binti Mohd Sani, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Azreen bin Azman, PhD
Senior Lecturer
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Datin Norwati Mustapha, PhD
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Aboubekeur Hamdi-Cherif, PhD
Professor
Qassim University
Saudi Arabia
(External Examiner)

RUSLI HAJI ABDULLAH, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 30 July 2018

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Nurfadhlina Binti Mohd Sharef, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Masrah Azrifah Binti Azmi Murad, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

Hamidah Binti Ibrahim, PhD
Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

____
ROBIAH BINTI YUNUS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: ________________________ Date: ___________________

Name and Matric No.: Hazrina Binti Sofian, GS39234

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: ______________________________

Name of Chairman of Supervisory Committee: Associate Professor Dr. Nurfadhlina Binti Mohd Sharef

Signature: ______________________________

Name of Member of Supervisory Committee: Associate Professor Dr. Masrah Azrifah Binti Azmi Murad

Signature: ______________________________

Name of Member of Supervisory Committee: Professor Dr. Hamidah Binti Ibrahim

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Answer type	The datatype that ANLD and ALID are expecting to retrieve from linked data
CEM	Component-based Evaluation Mechanism
KB	Knowledge base
KB terminology	Word in the knowledge base
LBSTM	Linguistic-based SPARQL Translation Model
Linguistic terminology	Word in natural language question posed by user
Linguistic triple pattern	Triple pattern that is constructed by SQA
LTI	Linguistic Model Triples Identification
MA	Meta-Mapping Aspect
MAPE-K Loop	Monitoring-Analysing-Planning-Executing over a shared Knowledge feedback loop
NER	Named Entity Recognition
NL	Natural Language
NLI	Natural Language Interface
POS Tag	Part-of-Speech Tagging
QALD	Question answering over linked data
RDF	Resource Description Framework
RDF triple pattern	Triple pattern that exist in linked data
QT	Query Triple
Resources	Multiple knowledge bases
SCNLP	Stanford CoreNLP
ALID	Adaptive Linked Data Disambiguation Model
SPARQL	Simple Protocol and RDF Query Language
SQA	Semantic Question Answering
ANLD	Adaptive based Natural Language Disambiguation Model
URI	Uniform Resource Identifier

CHAPTER 1

INTRODUCTION

1.1 Research Motivation

Linked data uses Resource Description Framework (RDF) that describes the relation between two data (Wood et al., 2014). The RDF data models consist of basic units known as a RDF triple patterns which consists of subject, predicate, and object. The subject and the object are the two data where the relation is described by the predicate. The data in the RDF format can be retrieved using Simple Protocol and the RDF Query Language (SPARQL) query. Uniform Resource Identifier (URI) provides a unique name for each RDF triple (DuCharme et al., 1997).

Semantic question answering (SQA) is the means to retrieve answer (Shekarpour et al., 2015) or formulate answers based on a natural language (NL) question (Hakimov et al., 2013) from the linked data. The NL question that is posed by users is translated to SPARQL query before executed to the linked data (Lopez et al., 2010). The answers are presented to the users in the NL or any other required forms. According to (Kaufmann & Bernstein, 2007), users prefer an SQA system that accepts an NL question to one that uses keywords, phrases, and a graphical interface. Furthermore, keyword-based has lower potential to automatically resolve word ambiguity (Gracia & Jorge, 2009) because as stated by Park et al., (2014) the keyword-based queries are lack a clear specification of the relations among words; which leads to users' inconvenience because they must perform additional tasks to convey the correct context for their question. This statement is supported by (Yahya et al., 2013), who found that SQA systems promote users' convenience by discovering relevant information in the linked data.

Based on review of literature that have been carried out (Hakimov et al., 2015; Shekarpour et al., 2015; Lopez et al., 2012; Kim & Cohen, 2013; Yahya et al., 2013A; Park et al., 2014; Yahya et al., 2013B; Lopez et al., 2007; Damljanovic et al., 2011; Hakimov et al., 2012; Habernal & Konopík, 2013; Hakimov et al., 2013; Vandic et al., 2012; Luo et al., 2014; Dong et al., 2017; Zafar et al., 2018; Khashabi et al., 2018; Dubey et al., 2018; Abdi et al., 2018; Gardner et al., 2018), this research concludes that the problems in SQA are the three disambiguations of the SQA: (1) disambiguate NL questions posed by users, (2) disambiguate linked environment such as aggregating data from across heterogeneous resources and heterogeneous domain to construct the SPARQL queries, and (3) disambiguate word ambiguity (Lopez et al., 2012; Shekarpour et al., 2015). The range of approaches have been developed and showed significant advances in handling NL questions while querying the linked data. Earlier literature shows that only simple NL question can be processed and rewritten into correct SPARQL query by involving the users to perform word disambiguation by choosing suitable concept and property names in the linked data (Damljanovic et al., 2011; Unger & Cimiano, 2011). This burdens the users and may construct wrong SPARQL queries especially when the users are not familiar with the structure of the linked data. Recently, there are very scarce existing word disambiguation studies for complex NL questions processing in the SQA. The result of an in depth study in the existing SQA system shows that the focus of researches in the SQA can be divided into two groups.

The first group of researcher is focusing on the problem in disambiguating the NL question formulations. However, this group does not focus on disambiguating linked data environment as they excludes all of the NL questions that involve complex linked data structure (i.e., SPARQL query with ORDER BY and FILTER) and also does not disambiguate word ambiguity (Hakimov et al., 2015; Dong et al., 2017; Abdi et al., 2018; Palangi et al., 2018). In the SQA context, the gap is that each user has their own style in writing the NL questions. Different style (ambiguity) in the NL question formulations which is unknown during the SQA development time may require the same answer from the linked data or may require a completely different answer from the linked data. The disambiguation of the NL question formulation can be divided into three meta-mapping aspects, namely: pattern variation in NL questions, complex NL questions (refer Table 2.5: The criteria of the complex NL question), and linguistic terminology which is the terms being used by user in formulating the NL question. An adaptive mechanism in natural language processing techniques is required as it has the capability to adapt an unknown NL question structure by using the disambiguation solution of the most similar question pattern that was previously learned (Refer Section 3.2.2). Other than that, new NL questions with variety of question patterns often interact with the SQA where these inputs shall be made used to bring SQA intelligence closer to reality. The SQA that learns new NL question patterns throughout their lifetimes and use such knowledge is beneficial for future decision making process.

The second group of researcher is focusing on disambiguating linked data environment and word disambiguation (Shekarpour et al., 2015; Yahya et al., 2013A; Park et al., 2014; Zafar et al., 2018; Khashabi et al., 2018; Dubey et al., 2018). However, this group does not focus on NL question disambiguation and the complex NL questions are disregarded in their evaluation. Moreover, there are also unknown NL question patterns which are unable to be processed in their evaluations.

In summary, there are three major disambiguation abilities of the SQA that are responsible for users' easy reliance to query the linked data. In like manner Kaufmann & Bernstein, (2007) stated that word disambiguation, the complexity of the SPARQL query construction and the lack of the SQA systems that are built for heterogeneous domains are the major limitations in the SQA system. However, there is no other researcher that works in combining the three disambiguations. These three disambiguations are needed to be combined because through empirical study that had been carried out, this research has found that the SPARQL components are determined by seven meta-mapping aspects. The seven meta-mapping aspects (MA) are: (MA1) pattern variation in NL question; (MA2) complex NL question; (MA3) linguistic terminologies; (MA4) data type variation (object/data properties); (MA5) heterogeneous resource referencing; (MA6) KB concept terminologies; (MA7) structure variation in the linked data. Meta-mapping aspects (1)-(3) are derived from the NL questions which are unknown during development time. These three meta-mapping aspects are disambiguated in Chapter 4. Meanwhile, meta-mapping aspects (4)-(7) derived from the linked data environment which is also unknown during development time. These four meta-mapping aspects are disambiguated in Chapter 5. Meta-mapping aspects (3) and (6) are linguistic terminologies and the KB concept terminologies. These two terminologies may require word disambiguation.

The seven meta-mapping aspects (MA) can exponentially form thousands of different merging meta-mapping scenarios because each user who has his/her own style of formulating NL questions and there are ever changing linked data environment to be considered in order to construct a SPARQL query.

$$\text{Meta Mapping Scenarios} = \sum \left(MA_1^a \times MA_2^b \times MA_3^c \times MA_4^d \times MA_5^e \times MA_6^f \times MA_7^g \right)$$

Each meta-mapping scenario shares the same SPARQL template. Thus, it is impossible to store thousands of rules or SPARQL templates to cater all meta-mapping scenarios which lead to the need for adaptive model to adapt SPARQL query based on the NL question and the linked data environment. This realisation motivates this research to incorporate this adaptive model into the design.

1.2 Research Problems

This section elaborates the research problems of this research. There are two problems of this research.

The first problem of this research is the unknown structure of NL questions poses by users which involves variation of patterns, multi-criteria of complexity and variation of linguistic terminologies. These three meta-mapping aspects need to be addressed simultaneously as the meta-mapping scenario determines the SPARQL component. There is a lack of SQA method that processes the variation of NL question patterns without manual modification, handling multiple criteria of complex NL questions (Shekarpour et al., 2015; Yahya et al., 2013A; Park et al., 2014) and addressing the selection of linguistic terminologies which distinguish different intentions of the users. For example, words can distinguish declarative sentences from interrogative sentences; the constructor selects such as “Show, Give or List” in the NL questions which indicates that the answer of this NL question returns as a list of results instead of one result (Ferre, 2012). Therefore, it is vital to analyse each linguistic terminology in an NL question before constructing the SPARQL query for answer retrieval.

The second problem of this research is the unknown of linked data environment which involves variation of datatype of the KB concept, the multiplicity resource referencing, the KB concept terminologies, and the variation linked data structure. These four meta-mapping aspects need to be addressed simultaneously as the meta-mapping scenario determines the SPARQL component. The multi-types of word ambiguity are also found in NL questions. The existing approaches are semi-automated as some SQA manually annotate mapping for SPARQL query construction (Makris et al., 2010). There also need to disambiguate multi types of word ambiguities which are related to linguistic terminologies and KB concept terminologies. Based on the interdisciplinary literature review, there are thirteen types of ambiguities that can occur in the SQA. The majority of the SQA that has been reviewed focused on resolving synonym ambiguity (Kim & Cohen, 2013; Yahya et al., 2013A; Vandic et al., 2012; Lopez et al., 2012; Lopez et al., 2010; Celikyilmaz, 2006; Hakimov et al., 2013; Yauri et al., 2013). Four SQA systems perform multiple types of word disambiguation. (Lopez et al., 2010) performs word disambiguation for synonym ambiguity, hypernym ambiguity and meronyms ambiguity using Semantic web (owl.sameAs). The remaining three SQAs perform word disambiguation for two types of ambiguity respectively. Unger & Cimiano (2011) perform word disambiguation for homonym ambiguity and synonym ambiguity using a method called Sortal Restriction and ontological reasoning. Kim & Baldwin (2013) perform word disambiguation for synonym and sentence segmentation using analogy-based interpretation. Yahya et al., (2013A) performs word disambiguation for synonym and sentence segmentation ambiguity using integer linear program. For example, one of the most recent researches in SQA (Hakimov et al., 2015) uses 54 lexical knowledges to manually disambiguate the conceptual ambiguity. Meanwhile, Shekarpour et al., (2015) reported on the failure to handle queries that require query expansions or query cleaning tasks.

1.3 Research Objectives

This section elaborates the research objectives of this research based on the above mentioned research problems. Two research objectives of this research are defined.

1. To propose an adaptive model to resolve the unknown structure of NL questions poses by users that requires simultaneous disambiguation for variations of question patterns, multi-criteria of NL question complexities and variations of linguistic terminologies (Chapter 4).

2. To propose an adaptive model to resolve unknown linked data environment that requires simultaneous disambiguation for the variations of datatype, resource heterogeneity, placement of subject-predicate-object in the triple pattern, the links between related KB concept and multiplicities of the RDF triple patterns. This adaptive model is also proposed to disambiguate multi types of word ambiguity (Chapter 5).

1.4 Research Scope

The formulated SQA conceptual framework is being integrated with an existing adaptive model to handle dynamic software requirements. The following describes the research scope for each of the SQA abilities.

The first SQA's ability is to understand variation forms of the NL questions posed by different users. In general, there are several types of sentences, for example declarative, imperative, interrogative, and exclamatory (Hooper, 1980). However, the SQA focuses only on imperative sentences and interrogative sentences because they are used by a user to ask questions. Both types of sentence can vary in length and complexity.

The second SQA's ability is to construct the SPARQL query in an uncertain linked data environment. In terms of structure variation in the linked data, this research handles homogeneous or heterogeneous resource referencing, single or multiple triple pattern(s), the placement of subject, predicate and object in each triple patterns. However, sensing the requirement to construct multiple triple patterns due to users' terminology that requires definition is not the scope of the research. For example, "Taikonout" by definition is "Astronaut from China". By using the word "Taikonout", adaptive model does not sense any un-used terminology given by the users. Thus, the adaptive model to construct the second triple patterns of an SPARQL query is not executed.

The third SQA's ability is to disambiguate word ambiguities. This research handles eight ambiguity types including new types of ambiguities, namely: homonym ambiguity, lexicon expansion ambiguity, lexicon reduction ambiguity, upper or lower case ambiguity, singular or plural ambiguity, semantic similarity/synonym ambiguity, space within noun phrase ambiguity, the underscored within noun phrase ambiguity. However, disambiguation is only to a certain extend, using semantic similarity matching between users' terminology and the linked data concept. The ambiguities due to definition and morphology are not within this research scope. There are also a few ambiguity types reported in literature but they are not tested using adaptive model because these types of ambiguity do not exist in **Q**uestion **A**nswering over **L**inked **D**ata (**QALD**) datasets: QALD-2, QALD-3, and QALD-4 test dataset.

Furthermore, there are other scenarios of superlative adjective and comparative adjective which are not within the scope of this research. This scenario occurs when there is a zero-matched between users' terminology and the KB concept. Thus, an SQA is required to retrieve answer based on the noun of the NL question for example "city". There is a need for an SQA to apply the SPARQL condition, ORDER BY, to sort the results, based on the size of the city, before selecting the result which is, also, not within the scope of this research.

On the other hand, superlative adjective and comparative adjective are able to be processed by this research if the adjective words exist in the linked data. For example, “Who was the first to climb Mount Everest?” and “How many inhabitants does the largest city in Canada have?” The word “first” and “largest”, are superlative adjective and this research is able to construct the SPARQL query using these complex words because the system able to retrieve KB concepts “firstAscentPerson” and the “largestCity” in the linked data.

1.5 Research Contributions

There are four contributions from this research which are:

The first contribution of this research is an Adaptive-based Natural Language Disambiguation (ANLD) Model that is integrated with the Linguistic-Based SPARQL Translation Model (LBSTM) for NL question disambiguation.

The second contribution of this research is an Adaptive-based Linked Data Disambiguation (ALID) Model that disambiguates the linked data environment that is ever-changing and growing. The ALID model provides the ability for an SQA to keep the initial SPARQL query and alter it, at the very least. This is performed by sensing specific SPARQL component/element that contradicts with the current linked data environment and alters that particular component/element using adaptive parameter control and component-based approach (refer Section 2.4.2). The LBSTM learns the new meta-mapping scenarios along with the new SPARQL templates for future use. Therefore, the same meta-mapping scenario in the future can be resolved by the ANLD model which is integrated with the LBSTM.

The third contribution is the ALID model that able to disambiguate the multi-types of word ambiguities without relying heavily on the lexicon knowledge.

The fourth contribution of this research is the Component-based Evaluation Mechanism (CEM). The CEM is capable to become a mean for other SQA researchers to identify the exact SPARQL component/element to be fixed. The CEM is also an evaluation mechanism in addition to the existing SQA evaluation framework that serves to eliminate bias result towards the SQA system which is capable to construct most of the SPARQL query parts accurately.

1.6 Organization of the Thesis

This section elaborates the remainder of the thesis.

The next chapter is divided into 2 major parts. The first part provides literature reviews regarding the existing SQA processes and the SQA conceptual framework which is proposed based on the literature. The complexity of the NL questions posed by users is thus, described. Then, the existing work on the word ambiguities in English and the identification of ambiguity types from multidisciplinary literature are summarized. Once the word ambiguity types are identified, the word ambiguity scenarios and the word ambiguity types that have been resolved in the SQA system are identified, too. Next, a review of the advanced disambiguation solutions for the SQA systems is presented, and finally, the second part provides literature review on adaptive and evolutionary, and in depth study on adaptive model and adaptive approaches (Chapter 2).

The research methodology is described in Chapter 3 while Chapter 4 discusses the formulation of the **A**daptive-based **N**atural **L**anguage **D**isambiguation (**ANLD**) model and results of evaluation on the ANLD model are discussed. Chapter 5 elaborates the new **C**omponent-based **E**valuation **M**echanism (**CEM**) to identify SPARQL components that are constructed by ANLD model that requires further enhancement. Then, the **A**daptive-based **LI**nked data **S**tructure **D**isambiguation (**ALID**) model is formulated and results of evaluation on the ALID model, which is too, are discussed. Finally, Chapter 6 presents the conclusion and potential avenues for future researches.

REFERENCES

Abdi, A., Idris, N., & Ahmad, Z. (2018). QAPD: An ontology-based question answering system in the physics domain. *Soft Computing, 22(1)*, 213–230.

Al-Khalifa, H. S., Al-Yahya, M. M., Bahanshal, A., & Al-Odah, I. (2009). SemQ: A proposed framework for representing semantic opposition in the Holy Quran using semantic web technologies. In *2009 International Conference on the Current Trends in Information Technology (CTIT)* (pp. 1–4). IEEE. http://doi.org/10.1109/CTIT.2009.5423145

Alsaeedan, W., & Menai, M. E. B. (2015). Self-adaptive genetic algorithm for the word sense disambiguation problem. In *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems* (Vol. 9101, pp. 581–590). http://doi.org/10.1007/978-3-319-19066-2

Arcaini, P., Riccobene, E., & Scandurra, P. (2015). Modeling and analyzing MAPE-K feedback loops for self-adaptation. In *Proceedings - 10th International Symposium on Software Engineering for Adaptive and Self-Managing Systems, SEAMS 2015* (pp. 13–23). IEEE Press. http://doi.org/10.1109/SEAMS.2015.10

Asher, N., & Denis, P. (2005). Lexical ambiguity as type disjunction. In *Third international Workshop on Generative Approaches to the Lexicon*.

Bear, J., & Hobbs, J. R. (1988). Localizing expression of ambiguity. In *Proceedings of the second conference on Applied natural language processing* (pp. 235–242). Morristown, NJ, USA: Association for Computational Linguistics. http://doi.org/10.3115/974235.974278

Bencomo, N., Grace, P., Flores, C., Hughes, D., & Blair, G. (2008). Genie: Supporting the model driven development of reflective, component-based adaptive systems. In *Proceedings of the 30th international conference on Software engineering* (pp. 811–814). http://doi.org/10.1145/1368088.1368207

Bhattacharyya, P. (2012). Natural language processing: A perspective from computation in presence of ambiguity, resource constraint and multilinguality. *CSI Journal of Computing, 1(2)*, 1–13.

Bresnan, J. (2003). Explaining morphosyntactic competition. In *The handbook of Contemporary Syntactic Theory*. Balti and Collins (eds.) Blackwell.

Brest, J., Greiner, S., Boskovic, B., Mernik, M., & Zumer, V. (2006). Self-adapting control parameters in differential evolution: A comparative study on numerical benchmark problems. *IEEE Transactions on Evolutionary Computation, 10(6)*, 646–657.

Brun, Y., Serugendo, G. D. M., Gacek, C., Giese, H., Kienle, H., Litoiu, M., ... & Shaw, M. (2009). Engineering self-adaptive systems through feedback Loops. Software engineering for self-adaptive systems. In *Software engineering for self-adaptive systems* (pp. 48–70). Springer, Berlin, Heidelberg.

Celikyilmaz, A. (2006). *A semantic question/answering system using Topic Models*. (Doctoral dissertation, PhD thesis, University of California, Berkeley).

Cheng, S. W., Garlan, D., & Schmerl, B. (2005). Making self-adaptation an engineering reality. In *Self-star properties in complex information systems* (Vol. 3460, pp. 158–173). Springer, Berlin, Heidelberg. http://doi.org/10.1007/11428589_11

Cimiano. H., & P. H. S. (2010). Knowledge engineering and management by the masses. In *International conference, EKAW 2010, Lisbon, Portugal, October 11-15, 2010; proceedings*. Springer Berlin Heidelberg.

Cleland, A. A., Gaskell, M. G., Quinlan, P. T., & Tamminen, J. (2006). Processing semantic ambiguity: Different loci for meanings and senses. In *Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 28, No. 28).*

Damljanovic, D., Agatonovic, M., & Cunningham, H. (2011). FREyA: An interactive way of querying linked data using natural language. In R. García-Castro, D. Fensel, & G. Antoniou (Eds.), *Extended Semantic Web Conference* (pp. 125–138). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-25953-1

Dixit, M., Upadhyay, N., & Silakari, S. (2015). An exhaustive survey on nature inspired optimization algorithms. *International Journal of Software Engineering and Its Applications*, *9(4)*, 91–104.

Dong, L., Mallinson, J., Reddy, S., & Lapata, M. (2017). Learning to paraphrase for question answering. In *arXiv preprint arXiv:1708.06022.*

Dubey, M., Banerjee, D., Chaudhuri, D., & Lehmann, J. (2018). EARL: Joint entity and relation linking for question answering over knowledge graphs. In *arXiv preprint arXiv:1801.03825.*

DuCharme, B. Laurent, S.S, Remenyi, D., Smith, S., dan White, T. (1997). *Learning SPARQL: Achieving maximum value from information systems: A process approach*. John Wiley & Sons, Inc.

DuCharme, B. (2013). *Learning SPARQL: Querying and updating with SPARQL 1.1*. O'Reilly Media, Inc.

Empson, W. (2004). *Seven types of ambiguity*. (Vol. 645). Random House.

Ferré, S. (2012). Squall: A controlled natural language for querying and updating rdf graphs. In *International Workshop on Controlled Natural Language* (pp. 11–25). Springer, Berlin, Heidelberg.

Ferreira, V. S., Slevc, L. R., & Rogers, E. S. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition, 96(3)*, 263–284.

Gardner, M., Grus, J., Neumann, M., Tafjord, O., Dasigi, P., Liu, N., ... & Zettlemoyer, L. (2018). AllenNLP: A deep semantic natural language processing platform. In *arXiv preprint arXiv:1803.07640.*

Garlan, D., Cheng, S.-W., Huang, A.-C., Schmerl, B., & Steenkiste, P. (2004). Rainbow: Architecture-based self adaptation with reusable infrastructure. *IEEE Computer*, *37*(10), 46–54. http://doi.org/10.1109/ICAC.2004.1301377

Gaver, W. W., Beaver, J., & Benford, S. (2003). Ambiguity as a resource for design. In *Proceedings of the SIGCHI conference on Human factors in computing systems. ACM* (pp. 233–240).

Geihs, K., Barone, P., Eliassen, F., Floch, J., Fricke, R., Gjorven, E., ... & Papadopoulos, G. A. (2009). A comprehensive solution for application level adaptation. *Software: Practice and Experience*, 385–422.

Gelbukh, A. (2011). Computational linguistics and intelligent text processing. In *7th international conference, cicling*. Springer.

Gillon, B. S. (1990). Ambiguity, generality, and indeterminacy: Tests and definitions. *Synthese*, *85*(3), 391–416. http://doi.org/10.1007/BF00484835

Girju, R., Badulescu, A., & Moldovan, D. (2006). Automatic discovery of part-whole relations. *Computational Linguistics, 32(1)*, 83–135.

Gleich, B., Creighton, O., & Kof, L. (2010). Ambiguity detection: Towards a tool explaining ambiguity sources. In *International Working Conference on Requirements Engineering: Foundation for Software Quality* (Vol. 6182, pp. 218–232). Springer Berlin Heidelberg.

Gracia, J., & Mena, E. (2009). Multiontology semantic disambiguation in unstructured web contexts. In *Proceedings of the 2009 K-CAP Workshop on Collective Knowledge Capturing and Representation* (pp. 1–9).

Grosz, B.J., Appelt, D.E., Martin, P.A., Pereira, F. C. N. . (1987). TEAM: An experiment in the design of transportable natural language interfaces. *Artificial Intelligence 32(2)*, 173 – 243.

Habernal, I., & KonopíK, M. (2013). SWSNL: semantic web search using natural language. *Expert Systems with Applications, 40(9)*, 3649–3664. http://doi.org/10.1016/j.eswa.2012.12.070

Hakimov, S., Oto, S. A., & Dogdu, E. (2012). Named entity recognition and disambiguation using linked data and graph based centrality scoring. In *Proceedings of the 4th international workshop on semantic web information management (p. 4). ACM.*

Hakimov, S., Tunc, H., Akimaliev, M., & Dogdu, E. (2013). Semantic question answering system over linked data using relational patterns. In *Proceedings of the Joint EDBT/ICDT 2013 Workshops . ACM* (pp. 83–88). New York, New York, USA: ACM Press. http://doi.org/10.1145/2457317.2457331

Hakimov, S., Unger, C., Walter, S., & Cimiano, P. (n.d.). Applying semantic parsing to question answering over linked data: Addressing the lexical gap. In *International Conference on Applications of Natural Language to Information Systems* (pp. 103–109). Springer, Cham.

Hamani, M. S., & Maamri, R. (2013). Word semantic similarity based on document's title. In *Database and Expert Systems Applications (DEXA), 2013 24th International Workshop* (pp. 43–47). IEEE.

He, H. (2011). *Self-adaptive systems for machine intelligence*. John Wiley & Sons.

Hooper, J. S. (1980). *A Quick English Reference*. University Press.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*(2), 137–194. http://doi.org/10.1016/S0364-0213(99)80005-6

Kaufmann, E., & Bernstein, A. (2007). How useful are natural language interfaces to the semantic web for casual end-users? In *The Semantic Web, 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference* (pp. 281–294).

Kaufmann, E., & Bernstein, A. (2010). Evaluating the usability of natural language query languages and interfaces to Semantic Web knowledge bases. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(4), 377-393.

Khakpour, N., Jalili, S., Talcott, C., Sirjani, M., & Mousavi, M. (2012). Formal modeling of evolving self-adaptive systems. *Science of Computer Programming*, *78*(1), 3–26. http://doi.org/10.1016/j.scico.2011.09.004

Khashabi, D., Sabharwal, T. K. A., & Roth, D. (2018). Question Answering as Global Reasoning over Semantic Abstractions.

Khawalda, M. I., & Al-Saidat, E. M. (2012). Structural ambiguity interpretation: A case study of arab learners of english. *Global Journal of Human-Social Science Research*, *12*(6).

Kim, J. D., & Cohen, K. B. (2013). Kim, J. D., & Cohen, K. B. (2013). Natural language query processing for SPARQL generation: A prototype system for SNOMED CT. In *Proceedings of BioLINK* (pp. 32–38).

Kim, S. N., & Baldwin, T. (2013). A lexical semantic approach to interpreting and bracketing English noun compounds. *Natural Language Engineering*, *19*(3), 385–407. http://doi.org/10.1017/S1351324913000107

Klepousniotou, E. (2002). The processing of lexical ambiguity: Homonymy and polysemy in the mental lexicon. *Brain and Language*, *81*(1–3), 205–223. http://doi.org/10.1006/brln.2001.2518

Lei, Y., Uren, V., & Motta, E. (2006). SemSearch: A search engine for the semantic web. In *International Conference on Knowledge Engineering and Knowledge Management* (Vol. 4248 LNAI, pp. 238–245). Springer, Berlin, Heidelberg.

Liddle, S. W. (2011). Model-driven software development. In *Handbook of Conceptual Modeling* (pp. 17–54). Springer, Berlin, Heidelberg.

Litoiu, M., Shaw, M., Tamura, G., Villegas, N. M., Müller, H. A., Giese, H., ... & Rutten, E. (2017). What can control theory teach us about assurances in self-adaptive software systems? In *Software Engineering for Self-Adaptive Systems III. Assurances* (pp. 90–134). Springer, Cham.

Lopez, V., Fernández, M., Motta, E., & Stieler, N. (2012). Poweraqua: Supporting users in querying and exploring the semantic web. *Semantic Web*, *3*(3), 249–265. http://doi.org/10.3233/SW-2011-0030

Lopez, V., Nikolov, A., Sabou, M., Uren, V., Motta, E., & d'Aquin, M. (2010). Scaling up question-answering to linked data. In *International Conference on Knowledge Engineering and Knowledge Management* (pp. 193–210). Springer Berlin Heidelberg.

Lopez, V., Uren, V., Motta, E., & Pasin, M. (2007). AquaLog: An ontology-driven question answering system for organizational semantic intranets. *Web Semantics: Science, Services and Agents on the World Wide Web*, *5*(2), 72–105. http://doi.org/10.1016/j.websem.2007.03.003

Luo, C., Liu, Y., Zhang, M., & Ma, S. (2014). Query ambiguity identification based on user behavior information. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8870*, 36–47.

Macías-Escrivá, F. D., Haber, R., Del Toro, R., & Hernandez, V. (2013). Self-adaptive systems: A survey of current approaches, research challenges and applications. *Expert Systems with Applications*, *40*(18), 7267–7279. http://doi.org/10.1016/j.eswa.2013.07.033

Makris, K., Gioldasis, N., Bikakis, N., & Christodoulakis, S. (2010). Ontology mapping and SPARQL rewriting for querying federated RDF data sources. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems* (pp. 1108–1117). Springer, Berlin, Heidelberg.

Mishra, A., & Misra, A. K. (2009). Component assessment and proactive model for support of dynamic integration in self adaptive system. *ACM SIGSOFT Software Engineering Notes*, *34*(4), 1–9. http://doi.org/10.1145/1543405.1543418

Naqvi, M. (2012). Claims and supporting evidence for self-adaptive systems–A literature review.

Palangi, H., Smolensky, P., He, X., & Deng, L. (2018). Question-answering with grammatically-interpretable representations. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence, New Orleans, LA.*

Pan, H. Y., & Felser, C. R. context effects in L. ambiguity resolution: E. from self-paced reading. (2011). Referential context effects in L2 ambiguity resolution: Evidence from self-paced reading. *Lingua*, *121*(2), 221–236. http://doi.org/10.1016/j.lingua.2010.08.003

Park, S., Shim, H., &Lee, G. G. (2014). ISOFT at QALD-4: Semantic similarity-based question answering system over linked data. In *CLEF*.

Pellerin, E., Pigeon, L., & Delisle, S. (2004). Self-adaptive parameters in genetic algorithms. In *Data Mining and Knowledge Discovery: Theory, Tools, and Technology VI* (p. Vol. 5433, pp. 53–65). International Society for Optics and Photonics. http://doi.org/10.1117/12.542156

Popentiu-Vladicescu, F., & Albeanu, G. (2016). Nature-inspired approaches in software faults identification and debugging. *Procedia Computer Science*, *92*, 6–12. http://doi.org/10.1016/j.procs.2016.07.315

Pradel, C., Haemmerle, O., & Hernandez, N. (2009). *A Natural Language Interface to Semantic Web using Modular Query Patterns A complete presentation of the SWIP*.

Quirk, R., Greenbaum, S., & Leech, G. & Svartvik, J. (1985). *A comprehensive grammar of the english language*. New York: London and New York: Longman.

Ravindranathan, M., & Leitch, R. (1998). Heterogeneous intelligent control systems. In *IEE Proceedings - Control Theory and Applications* (Vol. 145, p. 551). http://doi.org/10.1049/ip-cta:19982397

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal*

Behavior, *22*(3), 358–374. http://doi.org/10.1016/S0022-5371(83)90236-0

Rindflesch, T. C., & Aronson, A. R. (1994). Ambiguity resolution while mapping free text to the UMLS Metathesaurus. In *Proceedings of the Annual Symposium on Computer Application in Medical Care* (pp. 240–244). American Medical Informatics Association.

Ruckhaus, E., Baldizán, O., & Vidal, M. E. (2013). Analyzing Linked Data Quality with LiQuate. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"* (p. 629–638)). Springer, Berlin, Heidelberg.

Salehie, M., & Tahvildari, L. (2012). Towards a goal driven approach to action selection in self adaptive software. *Software: Practice and Experience*, *42(2)*, 211–233.

Schmeck, H., Müller-Schloer, C., Çakar, E., Mnif, M., & Richter, U. (2010). Adaptivity and self-organization in organic computing systems. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, *5*(3), 1–32. http://doi.org/10.1145/1837909.1837911

Shadbolt, N. R., Gibbins, N., Harris, S., & Glaser, H. (2004). CS AKTive space: representing computer science in the semantic web. In *Proceedings of the 13th international conference on World Wide Web* (pp. 384–392). ACM.

Sharef, N. M., & Mohd, S. A. (2012). Soft queries processing. In *Natural Language. International Conference*.

Shekarpour, S., Marx, E., Ngomo, A. C. N., & Auer, S. (2015). Sina: Semantic interpretation of user queries for question answering on interlinked data. *Journal of Web Semantics*, *30*, 39–51. http://doi.org/10.1016/j.websem.2014.06.002

Shekarpour, S., Ngonga Ngomo, A. C., & Auer, S. (2013). Question answering on interlinked data. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 1145–1156).

Snow, R., Jurafsky, D., & Ng, A. Y. (2004). Learning syntactic patterns for automatic hypernym discovery. In *Advances in Neural Information Processing Systems 17* (pp. 1297–1304).

Stageberg, N. C. (1968). Structural Ambiguity for English Teachers. In *Selected Addresses Delivered at the Conference on English Education* (p. No. 6, pp. 29–34). National Council of Teachers of English.

Steedman, M. (2000). *The Syntactic Process* ((Vol. 24).). Cambridge: MIT press.

Turnbull, J., Lea, D., Parkinson, D., Phillips, P., Francis, B., Webb, S., ... & Ashby, M. (2010). *Oxford advanced learner's dictionary* (8th ed.). Oxford University Express.

Unger, C., & Cimiano, P. (2011). Representing and resolving ambiguities in ontology-based question answering. In *Proceedings of the TextInfer 2011 Workshop on Textual Entailment. Association for Computational Linguistics.*

Vandic, D., Van Dam, J. W., & Frasincar, F. (2012). Faceted product search powered by the Semantic Web. *Decision Support Systems, 53*(3), 425–437. http://doi.org/10.1016/j.dss.2012.02.010

Vogel, T., Neumann, S., Hildebrandt, S., Giese, H., & Becker, B. (2009). Model-driven architectural monitoring and adaptation for autonomic systems. In *Proceedings of the 2009 International Conference on Autonomic Computing (ICAC'09)* (pp. 67–68). http://doi.org/10.1145/1555228.1555249

Weiss, G., Becker, K., Kamphausen, B., Radermacher, A., & Gerard, S. (2011). Model-driven development of self-describing components for self-adaptive distributed embedded systems. In *Software Engineering and Advanced Applications (SEAA), 2011 37th EUROMICRO Conference* (pp. 477–484). IEEE.

Weyns, D., & Georgeff, M. (2010). Self-adaptation using multi-agent systems. *Software, IEEE 27(1)*.

Wood D, Zaidman M, Ruth L, H. M. (2014). *Linked Data*. Manning Publications Co.

Xu, J., & Pottinger, R. (2014). Integrating domain heterogeneous data sources using decomposition aggregation queries. *Information Systems*, *39*(1), 80–107. http://doi.org/10.1016/j.is.2013.06.003

Yahya, M., Berberich, K., Elbassuoni, S., & Weikum, G. (2013). Robust question answering over the web of linked data. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management* (pp. 1107–1116). ACM.

Yahya, M., Berberich, K., Ramanath, M., & Weikum, G. (2013). On the SPOT: Question answering over temporally enhanced structured data. In *Proceedings of Workshop on Time-aware Information Access, TAIA2013.* Dublin, Ireland.

Yauri, A. R., Kadir, R. A., Azman, A., & Murad, M. A. A. (2013). Ontology semantic approach to extraction of knowledge from Holy Quran. In *2013 5th International Conference on Computer Science and Information Technology* (pp. 19–23). IEEE.

Yeom, K. (2015). Morphological approach for autonomous and adaptive system: The construction of three-dimensional artificial model based on self-reconfigurable modular agents. *Neurocomputing*, *148*, 100–111.

Zafar, H., Napolitano, G., & Lehmann, J. (2018). Formal query generation for question answering over knowledge bases. *ESWC*.

Zettlemoyer, L. S., & Collins, M. (2009). Learning context-dependent mappings from sentences to logical form. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2* (pp. 976–984). Association for Computational Linguistics.