A RULE-BASED APPROACH FOR CONVERTING WIKIPEDIA CONTENT INTO SEMANTIC RELATIONS

by

Nihal Yağmur Aydın

B.S., Computer Engineering, Middle East Technical University, 2013

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Computer Engineering

Boğaziçi University

2016

ACKNOWLEDGEMENTS

To my family and professors.

I would like to thank to my past professors, Prof. Christian Omlin, Assoc. Prof. Tolga Can and Prof. Athanasios Vasilakos for their encouragement to pursue graduate studies.

I am grateful to my advisor Prof. Tunga Güngör for his guidance and encouragement during my studies at Boğaziçi. Without his help, it would have been impossible to deepen my knowledge on natural language processing and finish my thesis.

I thank to Prof. Mehmet Ufuk Çağlayan for his encouragement to set this thesis topic, which I started to work on during my term project.

I also thank to my yoga teacher, Yogacharya, for his teachings on yoga. Without the techniques of yoga, it would not have been easy to focus on my work.

Lastly, I would like to thank my grandfather, who was my first teacher, teaching me to study in a planned way. I thank to my family members, my father, my mom and brother for their unconditional love and support.

ABSTRACT

A RULE-BASED APPROACH FOR CONVERTING WIKIPEDIA CONTENT INTO SEMANTIC RELATIONS

In this thesis, we propose a method for conversion from natural language into semantic relations. In this research, we focus on text written in highly unstructured form. The method is based on analysis of grammatical patterns of the sentences chosen from Wikipedia. Regular expressions are used for the generation of grammatical patterns. In addition to the grammatical structure of sentences, we also made use of the named entities to create semantic relations. Experiments on different types of relations showed that 71% and 82% success rates can be obtained for a threshold of 0.50 correctness rate.

ÖZET

VİKİPEDİ İÇERİĞİNİN ANLAMSAL İLİŞKİLERE DÖNÜŞÜMÜ İÇİN KURAL TABANLI BİR YAKLAŞIM

Bu tezde, doğal dilden anlamsal ilişkilere dönüşüm için bir yaklaşım ortaya koyuyoruz. Bu çalışmada, büyük oranda yapısal olmayan döküman içeriğine odaklandık. Yöntem, Vikipedi'den seçilen cümlelerin dilbilgisel model analizi üzerine temellenmiştir. Düzenli ifadeler, dilbilgisel modelin yaratılması için kullanılmıştır. Cümlelerin dilbilgisel yapısına ek olarak, anlamsal ilişkilerin yaratımı için adlı varlıklar kullanılmıştır. Farklı türlerdeki ilişkiler üzerinde yapılan deney sonuçlarına göre, 0.5 doğruluk payı üzerinden belirlenen sınıra göre %71 ve %82 oranında başarı görülmüştür.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

CD	Cardinal Number
IN	Proposition
JJ	Adjective
NER	Named Entity Recognition
NN	Noun (singular)
NNP	Proper Noun
NNS	Plural Noun
NP	Noun Phrase
POS	Part of Speech
PP	Propositional Phrase
VP	Verb Phrase

1. INTRODUCTION

The task of converting formal specification language into natural language could be easily done in these days. However, converting natural language into formal specification language is one of the tough one, having importance in formal verification and model checking. That task also concerns with the issues related to natural language understanding, artificial intelligence and question answering systems.

The problem of converting natural language has its roots from the year of 1977 (Bell et.al, 1977), where the issue was mostly focusing on converting software requirement into requirement system language (RSL).

There are currently studies consisting of retrieving formal specifications from structured text written in a proper format, such as getting formal verification properties from natural language documentation for HDL comments (Harris & Harris, 2015). However, such studies focus on conversion from sentences written in structured forms by using syntactical properties of sentences.

In this thesis, sentences written in natural language are chosen for converting text into semantic relations. In order to achieve this, grammatical patterns are analyzed in addition to outputs coming with Named Entity Recognition (NER). Semantic relations are created based on the outputs coming with Stanford NLP Parser and Stanford Named Entity Recognizer.

The novelty of that thesis comes from the fact that sentences being worked on are chosen from Wikipedia documents, where sentences are written in free format. In addition to that, patterns of grammatical structures are created from parse tree in order to derive relations and meaning from the text. The methods developed in this thesis solve the problem of natural understanding and question answering to extract meaning from the text written in natural language.

For the purpose of the thesis, Stanford NLP Parser is used to understand grammatical information of the sentences. Additionally, Stanford Named Entity Recognition tool is used to tag Person, Location and Organization information for the sentences. After that,

syntactical information and the results coming from NER are combined for creating relations.

1.1. Outline

This thesis is organized as follows: in the second Chapter, we present related work, in the third Chapter, we describe the methodology that is used. In Chapter 4, we explain our evaluation criteria and experiment results. Afterwards, in Chapter 5, we conclude the work and describe future work.

2. LITERATURE REVIEW

Several studies exist on the issue of conversion from natural language into formal specifications. Such conversions are usually done for model checking and also verification purposes. Examples of such studies are listed and examined in that section.

Researchers have worked on the issue of geographic information processing (Vessel U., 2002). Since we are also concerned with geographical content, that study has some relations with our thesis. In the study of Vessel *et al.* (2002) Geographical Information System (GIS) is supported with ontologies. Ontologies allow classification of knowledge in a proper and regulated way, whereas, in our study, we focus more on the relations derived from getting semantic relations from the sentences. Therefore, research on GIS made us consider the geographical properties.

In the research done by Harris I.G. (2013), context free grammar (CFG) is used to define rules for the syntactic parser. CFG is defined to capture English subset of interest, the parser generating the parse tree representation of a sentence. An example derivation of a CFG is shown as follows:

$$
\begin{aligned}
S &\rightarrow VP \\
VP &\rightarrow VB\ \ NN\ \ PP \\
VB &\rightarrow \text{“set”} \\
PP &\rightarrow IN\ \ NN \\
IN &\rightarrow \text{“to”} \\
NN &\rightarrow \text{“P”} \\
NN &\rightarrow \text{“one”}
\end{aligned}
$$

Figure 2.1. Derivation of CFG

Attribute grammars are also used in order to parse the class of assertions of sentences written in English producing System Verilog assertions, having semantic equivalence for English descriptions.

There are some studies consisting of retrieving formal specifications from structured text written in a proper format, such as getting formal verification properties from natural language documentation for HDL comments (Harris and Harris, 2015). However, such studies focus on conversion from sentences written in structured forms by using syntactical properties of sentences.

In another study, the conversion of natural language into semantic relations had been done by using outputs coming from Stanford Dependency Parser (Drechsler *et al*., 2014). SPARQL queries are run on the semantic relations of database. Some studies take conversion problem as a whole-sentence machine translation problem (Pust *et al.*, 2015). Translation issue had been separated into five steps: rule extraction, local feature extraction, language model calculation, decoding and tuning. In that study, Stanford Named Entity Recognizer is used as well for labeling information about location, person etc.

In a research done on knowledge representation (Zhang *et al.*, 2011), XML is examined where ontologies are combined to add semantics representation of knowledge, aiming to facilitate the development of web ontologies. In that approach, authors start with XML schema. Afterwards, they map elements of XML schema into OWL language. Some definitions in XML data become objects, some of them become new relations and some of them become attributes in that mapping. Lastly, reasoning tasks on XML schema are extended for ontologies.

In another study, (Lee and Bryant, 2002), two level grammars (TLG) are used in order to convert natural language to VDM++ specification. Input could be chosen as a data type, declaration, a rule, a statement for a rule or a meta sentence which contains information about the classifier or a set of rules. Knowledge base is translated into TLG and then into VDM++ specifications.

In embedded systems, ensuring correctness is of high importance. Therefore, operations regarding model checking are crucial. For that purpose, property checking is applied to address the issue by extracting properties from the specification in terms of temporal logic expressions which can be subsequently be checked by using algorithms known as model checker. Wordnet, Stanford Dependency Parser and UML are combined to

make model checking (Drechsler *et al.*, 2012).

Abstract Syntax Trees (ASTs) are generally used as intermediate representations in compiling code in a high level programming language to machine dependent code. In a study for extracting formal specification from natural language (Dinesh *et al.*), internal nodes of ASTs is chosen as operators (predicates), the subtrees that they dominate become the operands (arguments) and leaf nodes corresponds to variables or constants.

ARSENAL (Elenius *et al.*, 2014) is the system created which makes conversion and reasoning. In that system, relations coming with Stanford Dependency Parser are used for TTEthernet requirements document. Intermediate Representation (IR) table is made on a table due to events, numerals etc. ARSENAL first creates a graph from the given table. Each node is a mention entry and each (directed) edge indicates if a mention is related to other via relations. Linear Temporal Logic (LTL) is used for reasoning.

Conversion issue has also been used for deriving behavior specifications from textual use cases (Mencyl, 2004). Use case is first converted into a Pro-case by first constructing a finite automaton representing the use case. Afterwards, Pro-case is derived as a regular expression generating the same language as the automaton.

In another study focusing on formal verification of digital circuits using specifications expressed in English (Holt, 1999), Symbolic Model Verification(SMV) model checker program is used to get inferential information from the text written in computation tree logic (CTL). That system consists of four components: (i) a parser, (ii) a converter from semantic representations to CTL, (iii) the SMV model checker, and (iv) a module that mediates interaction between the three others, as well as handling input and output with the user.

Translation of natural language to OCL (Bajuva *et al.*, 2012) has also been concern where the input text is natural language specification of an OCL constraint for a UML class model. Sentence splitting, tokenization, POS tagging, lemmatization (morphological analysis) are the first steps of the conversion. Stanford Parser is used for conversion purposes. Sentences are translated into logical forms in the second step. Lastly, logical forms are mapped into OCL.

Mining text is closely related to extracting concepts from the document (Wang *et al.*,2008). For that purpose, in that study, concept maps are created and extracted on the experiments done on short texts. Generally, the subject of sentences represents the concept. Verbal phrase of the sentence is the object, representing a second concept. Relationship between subject and object are identified by the main verb in the sentence.

Parsing English into Abstract Meaning Representation (AMR) is a whole-sentence relation (May *et al.*, 2015). For the purpose of translation, the task had been chosen as the machine translation problem. By considering the conversion problem as a machine translation issue, translation phase had been separated into five steps: rule extraction, local feature extraction, language model calculation, decoding and tuning. In this study, Stanford Named Entity Recognizer is used for labeling information about location and person names, in order to support the representation.

3. METHODOLOGY

3.1. NLP Tools

There exist several tools for natural language processing. In this thesis, tools concerned with generating parse trees and named entity recognition are used.

3.1.1. Stanford Parser

Stanford Parser uses a lexicalized PCFG for parsing trees. It uses a top-down approach for parsing sentences. To find the most probable parse for grammatical structure (Klein & Manning, 2003) , A* parse is used to provide a proof for optimality.

In order to examine the parse tree, an example sentence is chosen as follows:

"Germany is a member of UN, NATO, the G8, the G20 and the OECD."

After that, sentence is put into Stanford Parser, resulting in the output as follows:

```
(ROOT
 (S
   (NP (NNP Germany))
   (VP (VBZ is)
     (NP
       (NP (DT a) (NN member))
  (PP (IN of)
        (NP
          (NP (NNP UN,))
          (VP (VBG NATO,)
            (S
              (NP (DT the) (NN G8,))
              (NP
                (NP (DT the) (NN G20))
                (CC and)
                (NP (DT the) (NNP OECD.)))))))))))
```

Figure 3.1. Stanford Parser

Tree structure which is shown above contains tags and words associated with it. Parse tree has the nodes containing words with grammatical tags.

If we try to reach nodes of the given parse tree we have the list of nodes as listed, where the numbers in the beginning shows the node number:

```
1 (ROOT (S (NP (NNP Germany)) (VP (VBZ is) (NP (NP (DT a) (NN member)) (PP (IN of) (NP (NP
(NNP UN) (NNP ) (NNP NATO) (NNP )) (SBAR (S (NP (DT the) (NN G8)) (VP (VBZ ) (NP (NP (DT
the) (NN G20)) (CC and) (NP (DT the) (NNP OECD)))))))))))

2 (S (NP (NNP Germany)) (VP (VBZ is) (NP (NP (DT a) (NN member)) (PP (IN of) (NP (NP (NNP UN)
(NNP ) (NNP NATO) (NNP )) (SBAR (S (NP (DT the) (NN G8)) (VP (VBZ ) (NP (NP (DT the) (NN
G20)) (CC and) (NP (DT the) (NNP OECD)))))))))))
3 (NP (NNP Germany))
4 (NNP Germany)
5 Germany
6 (VP (VBZ is) (NP (NP (DT a) (NN member)) (PP (IN of) (NP (NP (NNP UN) (NNP ) (NNP NATO)
(NNP )) (SBAR (S (NP (DT the) (NN G8)) (VP (VBZ ) (NP (NP (DT the) (NN G20)) (CC and) (NP (DT
the) (NNP OECD)))))))))))
7 (VBZ is)
8 is
9 (NP (NP (DT a) (NN member)) (PP (IN of) (NP (NP (NNP UN) (NNP ) (NNP NATO) (NNP )) (SBAR
(S (NP (DT the) (NN G8)) (VP (VBZ ) (NP (NP (DT the) (NN G20)) (CC and) (NP (DT the) (NNP
OECD)))))))))
10 (NP (DT a) (NN member))
11 (DT a)
12 a
13 (NN member)
14 member
15 (PP (IN of) (NP (NP (NNP UN) (NNP ) (NNP NATO) (NNP )) (SBAR (S (NP (DT the) (NN G8)) (VP
(VBZ ) (NP (NP (DT the) (NN G20)) (CC and) (NP (DT the) (NNP OECD)))))))))
16 (IN of)
17 of
18 (NP (NP (NNP UN) (NNP ) (NNP NATO) (NNP )) (SBAR (S (NP (DT the) (NN G8)) (VP (VBZ ) (NP
(NP (DT the) (NN G20)) (CC and) (NP (DT the) (NNP OECD)))))))
19 (NP (NNP UN) (NNP ) (NNP NATO) (NNP ))
20 (NNP UN)
21 UN
22 (NNP )
23
24 (NNP NATO)
25 NATO
26 (NNP )
27
28 (SBAR (S (NP (DT the) (NN G8)) (VP (VBZ ) (NP (NP (DT the) (NN G20)) (CC and) (NP (DT the)
(NNP OECD))))))
29 (S (NP (DT the) (NN G8)) (VP (VBZ ) (NP (NP (DT the) (NN G20)) (CC and) (NP (DT the) (NNP
 OECD)))))
30 (NP (DT the) (NN G8))
31 (DT the)
32 the
33 (NN G8)
34 G8
35 (VP (VBZ ) (NP (NP (DT the) (NN G20)) (CC and) (NP (DT the) (NNP OECD))))
36 (VBZ )
37
38 (NP (NP (DT the) (NN G20)) (CC and) (NP (DT the) (NNP OECD)))
```

Figure 3.2. Nodes of a Parse Tree

```
39 (NP (DT the) (NN G20))
40 (DT the)
41 the
42 (NN G20)
43 G20
44 (CC and)
45 and
46 (NP (DT the) (NNP OECD))
47 (DT the)
48 the
49 (NNP OECD)
50 OEC
```

Figure 3.2. Nodes of a Parse Tree (cont.)

As it can be understood from the nodes of the parse-tree in Figure 3.2, results are mixed and number of nodes of the tree cannot be estimated from the number of words in a given sentence. Because of that, it is required to use regular expressions to reach specific nodes of tree , supported by the patterns that can be generalized to set of sentences.

3.1.2. Regular Expressions for Parse Trees

Tregex is a utility for matching patterns in parse trees, which is based on relationships on trees and regular expressions on nodes associated with it. It uses regular expression library of Java. Owing to fact that this library has its own syntax for creating relations, explanations are required in order to provide more understanding for the patterns used in this thesis. TregexPattern is subclass of Tregex class, where matching of nodes are done based on the patterns specified by the programmer.

Here is the table for node-to-node matching library for TregexPattern (Tregex, Stanford):

Table 3.1. Tregex Pattern

Symbol	**Meaning**
A << B	A dominates B
A >> B	A is dominated by B
A < B	A immediately dominates B
A > B	A is immediately dominated by B
A $ B	A is a sister of B (and not equal to B)

Table 3.1. Tregex Pattern (cont.)

A .. B	A precedes B
A . B	A immediately dominates B
A ,, B	A follows B
A , B	A immediately follows B
A <<, B	B is a leftmost descendant of A
A <<- B	B is a rightmost descendant of A
A >>, B	A is a leftmost descendant of B
A >>- B	A is a rightmost descendant of B
A <, B	B is the first child of A
A >, B	A is the first child of B
A <- B	B is the last child of A
A >- B	A is the last child of B
A <` B	B is the last child of A
A >` B	A is the last child of B
A <i B	B is the ith child of A (i > 0)
A >i B	A is the ith child of B (i > 0)
A <-i B	B is the ith-to-last child of A (i > 0)
A >-i B	A is the ith-to-last child of B (i > 0)
A <: B	B is the only child of A
A >: B	A is the only child of B
A <<: B	A dominates B via an unbroken chain (length > 0) of unary local trees.
A >>: B	A is dominated by B via an unbroken chain (length > 0) of unary local trees.
A $++ B	A is a left sister of B (same as $.. for context-free trees)
A $-- B	A is a right sister of B (same as $,, for context-free trees)
A $+ B	A is the immediate left sister of B (same as $. for context-free trees)
A $- B	A is the immediate right sister of B (same as $, for context-free trees)
A $.. B	A is a sister of B and precedes B
A $,, B	A is a sister of B and follows B
A $. B	A is a sister of B and immediately precedes B
A $, B	A is a sister of B and immediately follows B
A <+(C) B	A dominates B via an unbroken chain of (zero or more) nodes matching description C
A >+(C) B	A is dominated by B via an unbroken chain of (zero or more) nodes matching description C
A .+(C) B	A precedes B via an unbroken chain of (zero or more) nodes matching description C
A ,+(C) B	A follows B via an unbroken chain of (zero or more) nodes matching description C

Table 3.1. Tregex Pattern (cont.)

A <<# B	B is a head of phrase A
A >># B	A is a head of phrase B
A <# B	B is the immediate head of phrase A
A ># B	A is the immediate head of phrase B
A == B	A and B are the same node
A <= B	A and B are the same node or A is the parent of B
A : B	[this is a pattern-segmenting operator that places no constraints on the relationship between A and B]
A <... { B ; C ; ... }	A has exactly B, C, etc as its subtree, with no other children.

Due to the Table 3.1, patterns are created by using grammatical tags of parse tree, such as "NP<CD".

3.1.3. Stanford Dependency Parser

Stanford Dependency Parser is used to define relations between words of a sentence.

Example of a sentence which is chosen for checking dependencies is: "Germany has a social market economy with a highly skilled labour force, a large capital stock, a low level of corruption, and a high level of innovation." After checking that sentence, dependency parser had given this output:

```
nsubj(has-2, Germany-1)
root(ROOT-0, has-2)
det(economy-6, a-3)
amod(economy-6, social-4)
compound(economy-6, market-5)
dobj(has-2, economy-6)
case(force-12, with-7)
det(force-12, a-8)
advmod(skilled-10, highly-9)
amod(force-12, skilled-10)
compound(force-12, labour-11)
nmod(economy-6, force-12)
det(stock-17, a-14)
amod(stock-17, large-15)
compound(stock-17, capital-16)
conj(economy-6, stock-17)
det(level-21, a-19)
amod(level-21, low-20)
appos(stock-17, level-21)
case(corruption-23, of-22)
nmod(level-21, corruption-23)
cc(economy-6, and-25)
```

Figure 3.3. Stanford Dependency Parser

```
det(level-28, a-26)
amod(level-28, high-27)
conj(economy-6, level-28)
case(innovation-30, of-29)
nmod(level-28, innovation-30)
```

Figure 3.3. Stanford Dependency Parser (cont.)

Outputs of dependency parser had been analyzed in order to generate rules, however, it had been understood that rules generated won't be helpful for creating relations.

3.1.4. Stanford Named Entity Recognizer

Stanford NER is used for labeling organizations, person and location information for sentences. It is also known as CRFC classifier, providing models for linear chain conditional random field (CRF) sequence models.

In order to illustrate Stanford NER, an example sentence is chosen:

"Germany is a member of UN, NATO, the G8, the G20 and the OECD."

Afterwards, output of NER is generated as follows:

"Germany/LOCATION is/O a/O member/O of/O UN/ORGANIZATION ,/O

NATO/ORGANIZATION ,/O the/O G8/O ,/O the/O G20/O and/O the/O

OECD/ORGANIZATION ./O"

By processing the generated output with NER tags, words could be separated from the given sentence by checking tags appended at the end of words.

3.2. Wikipedia Content

Wikipedia is a collaborative environment for information sharing and it is referred as an encyclopedia. On the other hand, since it is generated by humans, sentences can take variety of forms like every-day language usage. In this thesis, contents related to countries are selected from Wikipedia pages.

Set of Sentences chosen for the country “Germany” from Wikipedia:

1- Germany, officially the Federal Republic of Germany is a federal parliamentary republic in West-Central Europe.
2- It includes 16 constituent states and covers an area of 357,021 square kilometers (137,847 sq mi) with a largely temperate seasonal climate.
3-With about 81.5 million inhabitants, Germany is the most populous member state in the European Union.
4-Germany is a member of the United Nations, NATO, the G8, the G20, and the OECD.
5-Most of Germany has a temperate seasonal climate dominated by humid westerly winds.
6-Germany is a federal, parliamentary, representative democratic republic.
7-Germany has a civil law system based on Roman law with some references to Germanic law.
8-Germany has a social market economy with a highly skilled labour force, a large capital stock, a low level of corruption, and a high level of innovation.
9-According to the 2011 German Census, Christianity is the largest religion in Germany, claiming 66.8% of the total population.
10-Albrecht Dürer, Hans Holbein the Younger, Matthias Grünewald and Lucas Cranach the Elder were important German artists of the Renaissance, Peter Paul Rubens and Johann Baptist Zimmermann of the Baroque, Caspar David Friedrich and Carl Spitzweg of Romanticism, Max Liebermann of Impressionism and Max Ernst of Surrealism.
11-German literature can be traced back to the Middle Ages and the works of writers such as Walther von der Vogelweide and Wolfram von Eschenbach. Well-known German authors include Johann Wolfgang von Goethe, Friedrich Schiller, Gotthold Ephraim Lessing and Theodor Fontane.
12-Richard Strauss was a leading composer of the late Romantic and early modern eras. Karlheinz Stockhausen and Hans Zimmer are important composers of the 20th and early 21st centuries.
13-Albrecht Dürer, Hans Holbein the Younger, Matthias Grünewald and Lucas Cranach the Elder were important German artists of the Renaissance, Peter Paul Rubens and Johann Baptist Zimmermann of the Baroque, Caspar David Friedrich and Carl Spitzweg of Romanticism, Max Liebermann of Impressionism and Max Ernst of Surrealism.
14-Well-known international brands include Mercedes-Benz, BMW, SAP, Volkswagen, Audi, Siemens, Allianz, Adidas, Porsche, and DHL.
15- It lies mostly between latitudes 47° and 55° N and longitudes 5° and 16° E.
16- Notable German physicists before the 20th century include Hermann von Helmholtz, Joseph von Fraunhofer and Gabriel Daniel Fahrenheit, among others.
17- Numerous mathematicians were born in Germany, including Carl Friedrich Gauss, David Hilbert, Bernhard Riemann, Gottfried Leibniz, Karl Weierstrass, Hermann Weyl and Felix Klein.

Figure 3.4. Sentences chosen from Wikipedia

As it is illustrated in Figure 3.4, sentences are chosen due to information they contain. The ones having words starting with “it” are changed to make the sentence contain country name, replacing it with the related county name in the document.

3.2.1. Sentence Selection and Semantic Relations

After analysis is done on Wikipedia pages of countries, we chose important information about the countries. In order to synthesize knowledge regarding countries, it had been essential to have the name of the country inside the sentence. Moreover, sentences are chosen due to information they contain. Sentences having information regarding area, population, religion, climate, geographical coordinates, republic, state, economy, neighboring countries and membership for organizations are chosen for countries.

Additionally, general information regarding scientists, mathematicians, artists and brand names are also selected from the set of sentences.

Table 3.2. Relations

Relation	**Example Sentence**
border	Turkey is bordered by eight countries: Syria and Iraq to the south; Iran, Armenia, and the Azerbaijani exclave of Nakhchivan to the east; Georgia to the northeast; Bulgaria to the northwest; and Greece to the west.
area	With a territory of 110,994 square kilometers (42,855 sq mi), Bulgaria is Europe's 16th-largest country.
member	Germany is a member of UN, NATO, the G8, the G20 and the OECD.
population	The Netherlands had an estimated population of 16,785,403 on 30 April 2013.
location	Italy is a unitary parliamentary republic in Europe.
economy	The Netherlands has a market-based mixed economy, ranking 17th of 177 countries according to the Index of Economic Freedom.
republic	Bulgaria is a unitary parliamentary republic with a high degree of political, administrative, and economic centralisation.
state	Russia is a sovereign state in northern Eurasia.
religion	Christianity is currently the largest religion in the Netherlands, accounting for about one-third of the population.
climate	Ukraine has a mostly temperate continental climate, although the southern coast has a humid subtropical climate.
geo	Egypt lies primarily between latitudes 22° and 32°N, and longitudes 25° and 35°E.
author	Well-known German authors include Johann Wolfgang von Goethe, Friedrich Schiller, Gotthold Ephraim Lessing and Theodor Fontane.
composer	In the 19th century the most popular composers were: Józef Elsner and his pupils Fryderyk Chopin and Ignacy Dobrzyński.
artist	Distinguished contemporary artists include Roman Opałka, Leon Tarasewicz, Jerzy Nowosielski, Wojciech Siudmak, Mirosław Bałka, and Katarzyna Kozyra and Zbigniew Wąsiel in the younger generation.

Table 3.2. Relations (cont.)

physicist	Notable German physicists before the 20th century include Hermann von Helmholtz, Joseph von Fraunhofer and Gabriel Daniel Fahrenheit, among others.
mathematician	Numerous mathematicians were born in Germany, including Carl Friedrich Gauss, David Hilbert, Bernhard Riemann, Gottfried Leibniz, Karl Weierstrass, Hermann Weyl and Felix Klein.
brand	The new car market is dominated by domestic brands such as Renault (27% of cars sold in France in 2003), Peugeot (20.1%) and Citroën (13.5%).

As it is described on Table 3.2, relations are chosen based on the knowledge that can be retrieved from sentences, after analyzing sentences in Wikipedia pages. After identifying relation names, rule generation is required to be done. For that purpose, abstract representation of chosen relations are considered, taking into account of current tools and libraries of natural language processing.

Abstract representation of relations are created after the analysis of sentences as follows:

(i) relation_name (country_name, property).

(ii) relation_name (property).

3.3. Proposed Method

In the method proposed for the solution of generating relations, syntactical and semantical properties of sentences are considered. In order to have semantic relations, there had been a need to generate rules which are valid for different sentences. Therefore, proposed method consists of a rule-based approach.

3.3.1. Syntactical Properties

3.3.1.1. Parse Trees. The Stanford Parser example of a sentence “Bulgaria is a unitary parliamentary republic with a high degree of political, administrative, and economic centralisation.” is as follows:

```
Output:

(ROOT
  (S
    (NP (NNP Bulgaria))
    (VP (VBZ is)
      (NP
        (NP (DT a) (JJ unitary) (JJ parliamentary) (NN republic))
        (PP (IN with)
          (NP
            (NP (DT a) (JJ high) (NN degree))
            (PP (IN of)
              (NP
                (NP (JJ political))
                (, ,)
                (NP (JJ administrative))
                (, ,)
                (CC and)
                (NP (JJ economic) (NN centralisation))))))))
    (. .)))
```

Figure 3.5. Parse Tree

As it can be seen on the Figure 3.5 above, Stanford Parser gives a parse tree having tags, which can be inside of each other. Due to the fact that parse-tree contains lots of nodes which are connected to each other, in order to get the desired information and eliminate unnecessary information, it is required to separate group of words based on grammatical structure of sentences. Therefore, there had been a need to identify patterns for creation of relations amongst sentences.

3.3.1.2. Grammatical Patterns. Patterns are needed to generate rules in sentences so that relations are created. In order to achieve this, firstly, sentences are parsed and divided into parts as follows:

Table 3.3. Breaking the sentences into three parts

NP	**VP**	**NP**
(NP (NNP Germany))	(VP (VBZ is)	(NP (DT a) (NN member)) (PP (IN of) (NP (NP (DT the) (NNP United)) (VP (VBG Nations,NATO,) (S (NP (DT the) (NN G8,)) (NP

Table 3.3. Breaking the sentences into three parts (cont.)

		(NP (DT the) (NN G20,)) (CC and) (NP (DT the) (NNP OECD.))))))))))
(NP (DT The) (NNPS Netherlands))	(VP (VBZ is) (VP (VBN described)	(PP (IN as) (NP (DT a) (JJ constitutional) (NN state)))))
(NP (NNP Bulgaria))	(VP (VBZ is)	(NP (DT a) (JJ parliamentary) (NN democracy)) (SBAR (WHPP (IN in) (WHNP (WDT which))) (S (NP (DT the) (ADJP (RBS most) (JJ powerful)) (JJ executive) (NN position)) (VP (VBZ is) (NP (NP (DT that)) (PP (IN of) (NP (JJ prime) (NN minister))))))))
(NP (NNP France))	(VP (VBZ has)	(NP (NP (DT a) (JJ mixed) (NN economy)) (SBAR (WHNP (WDT that)) (S (VP (VBZ combines) (NP (JJ extensive) (JJ private) (NN enterprise)) (PP (IN with) (NP (JJ substantial) (NN state) (NN enterprise) (CC and) (NN government) (NN intervention)))))))

Partitioning the sentences into different parts had been first step to analyze the grammatical structure of the sentences; so that it had been possible to look them in detail. When the sentences are broken into parts as NP-VP-NP, it had been easier to study the structures as patterns. It is found that important information regarding sentences could be found inside of noun phrases.

Due to that fact, patterns inside of noun phrases are examined in more detail as illustrated on Table 3.4.

Table 3.4. Details of First Noun Phrase

1st noun phrase	**NNP / NNPS**	**JJ- NN/NNS,NN/NNS**
(NNP Germany)	(NNP Germany)	
(DT The) (NNPS Netherlands)	(NNPS Netherlands)	
(NNP France)	(NNP France)	
(NNP Bulgaria)	(NNP Bulgaria)	

As it can be seen on Table 3.4, sentences are chosen based on the knowledge in first noun phrases, coming before verbs, which usually contains proper nouns of country names. That had been basis of creation of relations from sentences.

Table 3.5. Details of Second Noun Phrase

2nd noun phrase	**NNP**	**JJ-NNS/NN**
(NP (DT a) (NN member)) (PP (IN of) (NP (NP (DT the) (NNP United)) (VP (VBG Nations,NATO,) (S (NP (DT the) (NN G8,)) (NP (NP (DT the) (NN G20,)) (CC and) (NP (DT the) (NNP OECD.))))))))))	(NNP OECD.)	(NN G20,) (NN G8,)
(PP (IN as) (NP (DT a) (JJ constitutional) (NN state)))))		((JJ constitutional) (NN state))
(NP (DT a) (JJ parliamentary) (NN democracy)) (SBAR (WHPP (IN in) (WHNP (WDT which))) (S (NP (DT the) (ADJP (RBS most) (JJ powerful)) (JJ executive) (NN position)) (VP (VBZ is) (NP (NP (DT that)) (PP (IN of) (NP (JJ prime) (NN minister))))))))		(JJ parliamentary) (NN democracy) (JJ executive) (NN position) (JJ prime) (NN minister)

By analyzing patterns listed above for the second noun phrase, nouns, proper nouns and adjectives are separated so that creation of relations had been possible.

(i) Pattern for Area Relation: To illustrate the idea of area, the example sentence is chosen: "Italy covers an area of 301,338 km2 (116,347 sq mi) and has a largely temperate seasonal climate; due to its shape, it is often referred to in Italy as lo Stivale (the Boot)." According to that example, the given sentence had been parsed as follows:

```
(ROOT
 (S
  (NP (NNP Italy))
  (VP
   (VP (VBZ covers)
    (NP
     (NP (DT an) (NN area))
     (PP (IN of)
      (NP (CD 301,338) (JJ km2) (CD (116,347) (NN sq) (NNS mi))))))
   (CC and)
   (VP (VBZ has)
    (VP
     (NP (DT a)
      (ADJP (RB largely) (JJ temperate))
      (JJ seasonal) (NN climate;))
     (ADJP (JJ due)
      (PP (TO to)
       (NP
        (NP (PRP$ its) (NN shape,))
        (SBAR
         (S
          (NP (PRP it))
          (VP (VBZ is)
           (ADVP (RB often))
           (VP (VBN referred)
            (S
             (VP (TO to)
              (VP
               (PP (IN in)
                (NP (NNP Italy)))
               (PP (IN as)
                (NP (NNP lo) (NNP Stivale) (NNP (the) (NNP Boot).))))))))))))))))
```

Figure 3.6. Pattern for Area Relation

Afterwards, for the generation of "area" relation, pattern of NP < CD is extracted for the first child of parse tree, returning the result as:

(NP (CD **301,338**) (JJ km2) (CD (116,347) (NN sq) (NNS mi))).

Following that, first cardinal number, which is (CD 301,338) is extracted, having

301,338 as the output to be put into area relation.

(ii) Pattern for Population Relation: To illustrate the idea of population relation, the following sentence is chosen: "The population of Bulgaria is 7,364,570 people according to the 2011 national census." giving the parse tree as follows:

```
(ROOT
  (S
    (NP
      (NP (DT The) (NN population))
      (PP (IN of)
        (NP (NNP Bulgaria))))
    (VP (VBZ is)
      (NP (CD 7,364,570) (NNS people))
      (PP (VBG according)
        (PP (TO to)
          (NP (DT the) (CD 2011) (JJ national) (NN census.)))))))
```

Figure 3.7. Pattern for Population Relation

Afterwards, in order to generate population relation, NP < CD pattern is extracted from the parse tree, getting the first child as follows: (NP (CD **7,364,570**) (NNS people)). Following that, number associated with CD tag is extracted for the creation of population relation.

(iii) Pattern for Republic Relation: To illustrate the idea of creation of a relation for republic, following sentence is chosen: "Bulgaria is a unitary parliamentary republic with a high degree of political, administrative, and economic centralisation." After the sentence is created, it is parsed as follows:

```
(ROOT
  (S
    (NP (NNP Bulgaria))
    (VP (VBZ is)
      (ADVP (RB a unitary))
      (NP
        (NP (JJ parliamentary) (NN republic))
        (PP (IN with)
          (NP
            (NP (DT a) (JJ high) (NN degree))
            (PP (IN of)
(NP                (NP (JJ political) (JJ ) (JJ administrative) (NNS ))
                  (CC and)
                  (NP (JJ economic) (NN centralisation)))))))))))
```

Figure 3.8. Pattern for Republic Relation

Afterwards, in order to get republic relation, NP < JJ & <<NN is extracted from the parse tree, getting first child having that pattern. Result of that compilation results in the output "(NP (JJ **parliamentary**) (NN **republic**))". Following that, words associated with adjective (JJ) and noun (NN) are extracted from the compiled pattern so that it had been possible to generate a semantic relation.

(iv) Pattern for Economy Relation: To illustrate the logic behind economy relation, following sentence is chosen: "The Netherlands has a market-based mixed economy, ranking 17th of 177 countries according to the Index of Economic Freedom." Following that, parse tree is generated as follows:

```
(ROOT
  (S
    (NP (DT The) (NNPS Netherlands))
    (VP (VBZ has)
      (NP
        (NP (DT a) (JJ market-based) (JJ mixed) (NN economy))
        (VP (VBG )
          (NP
            (NP (JJ ranking) (NNS 17th))
            (PP (IN of)
              (NP (CD 177) (NNS countries))))
          (PP (VBG according)
            (PP (TO to)
              (NP
                (NP (NN the Index))
                (PP (IN of)
                  (NP (NNP Economic) (NNP Freedom.)))))))))))
```

Figure 3.9. Pattern for Economy Relation

Afterwards, in order to get economy relation, NP < JJ & <<NN is extracted from the parse tree, getting the first child having that pattern. Result of that compilation results in the output "(NP (DT a) (JJ **market-based**) (JJ mixed) (NN **economy**))" . Following that, words associated with adjective (JJ) and noun (NN) are extracted from the compiled pattern so that it had been possible to generate a semantic relation.

(v) Pattern for Geo Relation: To illustrate the logic behind "geo" relation, the following sentence is chosen:

"Austria lies between latitudes 46° and 49° N, and longitudes 9° and 18° E."

After that, the sentence is parsed as illustrated on Figure 3.10.

```
(ROOT

 (S
  (NP (NNP Austria))
  (VP (VBZ lies)
   (PP (IN between)
    (NP
     (NP
      (ADJP (JJ latitudes)
       (NP (CD 46°)
        (CC and)
        (CD 49°)))
      (NNP N) (NNP ))
     (CC and)
     (NP
      (ADJP (JJ longitudes)
       (NP (CD 9°)
        (CC and)
        (CD 18°)))
      (NN E)))))))
```

Figure 3.10. Pattern for Geo Relation

The sentences having the words latitude and longitude are chosen for geographic coordinates. Afterwards, cardinal numbers (CD) are separated so that creations of the relations had been possible.

(vi) Pattern for Religion Relation: To illustrate the logic behind "religion" relation, the following sentence is chosen:
"Christianity is currently the largest religion in the Netherlands, accounting for about one-third of the population." After that, the sentence is parsed as follows:

```
(NP

 (NP (NNS Christianity is))
 (NP
  (NP (RB currently) (DT the) (JJS largest) (NN religion))
  (PP (IN in)
   (NP
    (NP (DT the) (NNP Netherlands) (NNP ) (NN accounting))
    (PP (IN for)
     (NP
      (NP
       (QP (RB about) (NN one-third)))
      (PP (IN of)
       (NP (DT the) (NN population.)))))))))
```

Figure 3.11. Pattern for Religion Relation

Afterwards, for the relation of "religion", following pattern is compiled: "NP < NP", making it possible to get the output "(NNS Christianity is)" for the creation of the

relation. As it can be understood from that example, problems in Stanford Parser also affect the results of the relation.

(vii) Pattern for Member Relation: To illustrate the idea behind the creation of member relation, the following sentence is chosen: "Spain is a member of UN, EU,CoE, OEI , NATO, OECD, WTO and many other international organisations." Following that, the sentence is parsed as follows:

```
(ROOT
 (S
  (NP (NNP Spain))
  (VP (VBZ is)
   (NP
    (NP (DT a) (NN member))
    (PP (IN of)
     (NP
      (NP (NNP UN,) (NNP EU,CoE,) (NNP OEI))
      (, ,)
      (NP (NNP NATO,) (NNP OECD,) (NNP WTO))
      (CC and)
      (NP (JJ many) (JJ other) (JJ international) (NN organisations.))))))))
```

Figure 3.12. Pattern for Member Relation

According to the output of the parse tree, it has been found that there is no grammatical pattern for the relation; whereas it is more based on Named Entity Recognition, which will be described in Chapter 3.3.2.

(viii) To illustrate the idea behind the climate relation, following sentence is chosen:

"Ukraine has a mostly temperate continental climate, although the southern coast has a humid subtropical climate." The parse tree generated after the selection of a sentence is as follows:

```
(ROOT
 (S
  (NP (NNP Ukraine))
  (VP (VBZ has)
   (NP
    (NP (DT a) (JJ mostly temperate continental) (NN climate) (NN ))
    (SBAR (IN although)
     (S
      (NP (DT the) (JJ southern) (NN coast))
      (VP (VBZ has)
       (VP (VBN a humid)
        (NP (JJ subtropical) (NN climate.)))))))))
```

Figure 3.13. Pattern for Climate Relation

Afterwards, for the relation of "climate", following pattern is compiled: " NP < JJ & <<NN", making it possible to get the output "(NP (DT a) (JJ **mostly temperate continental**) (NN **climate**) (NN))" for the creation of the relation.

(ix) Pattern for State Relation: To illustrate the idea behind the state relation, following sentence is chosen: "Belgium is a sovereign state in Western Europe." Following that, the parse tree is generated as follows:

```
(ROOT

 (S
  (NP (NNP Belgium) (NNP ))
  (VP (VBZ is)
   (NP
    (NP (DT a) (JJ sovereign) (NN state))
    (PP (IN in)
     (NP (NNP Western) (NNP Europe.)))))))
```

Figure 3.14. Pattern for State Relation

Afterwards, for the relation of "state", following pattern is compiled: " NP < JJ & <<NN", getting the first child of tree; making it possible to get the output "(NP (DT a) (JJ sovereign) (NN state))." After the separation of tags and words from the output, relations could be created.

(x) Pattern for Location Relation: To illustrate the idea behind the location relation, the following sentence is chosen: "Italy is a unitary parliamentary republic in Europe." Following that, the parse tree is generated as follows:

```
(ROOT

 (S
  (NP (NNP Italy))
  (VP (VBZ is)
   (NP
    (NP (DT a) (JJ unitary) (JJ parliamentary) (NN republic))
    (PP (IN in)
     (NP (NNP Europe.)))))))
```

Figure 3.15. Pattern for Location Relation

Afterwards, following pattern is compiled: "PP < NP," getting first child of the tree; making it possible to get the output **"**PP (IN in) (NP (NNP **Europe**.)))". After the separation of tags and words, relations could be created.

(xi) Pattern for Border Relation: To illustrate the idea behind the border relation, the following sentence is chosen: "Turkey is bordered by eight countries:

Syria and Iraq to the south; Iran, Armenia, and the Azerbaijani exclave of Nakhchivan to the east; Georgia to the northeast; Bulgaria to the northwest; and Greece to the west.” Following that, the parse tree is generated as follows:

```
(ROOT

 (S
  (NP (NNP Turkey))
  (VP (VBZ is)
   (VP (VBN bordered)
    (PP (IN by)
     (NP
      (NP (CD eight) (NNP countries:) (NNP Syria))
      (CC and)
      (NP (NNP Iraq))))
    (PP
     (PP (TO to)
      (NP
       (NP (DT the) (NNP south;) (NNP Iran,) (NNP Armenia,))
       (CC and)
       (NP
        (NP (DT the) (NNP Azerbaijani) (NN exclave))
        (PP (IN of)
         (NP (NNP Nakhchivan))))))
(PP (TO to)
(NP (DT the) (NNP east;) (NNP Georgia)))

  (PP (TO to)
      (NP
       (NP (DT the) (NNP northeast;) (NNP Bulgaria))
       (PP (TO to)
        (NP (DT the) (NNP northwest;)
         (CC and)
         (NNP Greece))))))
    (PP (TO to)
     (NP (DT the) (NN west.)))))))
```

Figure 3.16. Pattern for Border Relation

Afterwards, the parsed tree is examined. There had been no grammatical pattern for the relation; whereas it is more based on Named Entity Recognition.

3.3.2. Semantic Properties

Semantic properties are related to obtaining meaning from sentences. In order to create relations from sentences, some of the words needed to be identified firstly in terms of their meaning. By finding searched words for relations and benefiting from NER, it has been possible to derive meaning semantically from the sentences.

3.3.2.1. Named Entity Recognition (NER). In that study, NER had been helpful for detecting country names for getting country-specific information from the sentences. In addition to that, organization and person names could be identified by benefiting from NER.

- **NER for LOCATION Names**

Identifying word's meaning as location had been beneficial to create relations. For instance, for the sentence: “Turkey is bordered by eight countries: Syria and Iraq to the south; Iran, Armenia, and the Azerbaijani exclave of Nakhchivan to the east; Georgia to the northeast; Bulgaria to the northwest; and Greece to the west.”, words having location tag had been identified and listed as follows: [Turkey, Syria, Iraq, Iran,/O, Armenia,/O, Georgia, Bulgaria, Greece]

- **NER for ORGANIZATION Names**

Organization names had played a crucial role for member and brand relations, so that creating relations had been possible. As an example, for the sentence: “Spain is a member of UN, EU,CoE, OEI , NATO, OECD, WTO and many other international organisations.” , organization names are listed as follows: [UN , EU CoE/O , NATO , OECD , WTO].

- **NER for PERSON Names**

Person names had been important for relations having one variable, such as author, composer, artist, mathematician and physicist. As an example, for the sentence: “Distinguished contemporary artists include Roman Opałka, Leon Tarasewicz, Jerzy Nowosielski, Wojciech Siudmak, Mirosław Bałka, and Katarzyna Kozyra and Zbigniew Wąsiel in the younger generation.”

Due to input mentioned above, output for Person names had been: [Opałka, Leon, Tarasewicz, Jerzy, Nowosielski, Wojciech, Siudmak, Mirosław, Bałka, Katarzyna, Kozyra, Zbigniew, Wąsiel]

3.3.3. Combination of Semantic and Syntactical Properties

It had been possible to create relations by combining semantic and syntactical properties of the sentences, which are described in more detailed in that section.

(i) Area Relation: For the creation of area relation, following pattern is created: relation_name(LOCATION, CD). Afterwards, the example sentence is chosen as follows: "Italy covers an area of 301,338 km2 (116,347 sq mi) and has a largely temperate seasonal climate; due to its shape, it is often referred to in Italy as lo Stivale (the Boot)." The result of that input had been:

area(Italy, 301,338)

(ii) Population Relation: For the creation of population relation, following pattern is created: relation_name (LOCATION, CD). Afterwards, example sentence is chosen as follows: "The population of Bulgaria is 7,364,570 people according to the 2011 national census." Following that, the result of that input had been:

population(Bulgaria, 7,364,570)

(iii) Republic Relation: For the creation of republic relation, following pattern is created: relation_name (LOCATION, JJ). Following that, an example sentence is chosen: "Bulgaria is a unitary parliamentary republic with a high degree of political, administrative, and economic centralisation." Therefore, the result of input had been:

republic(Bulgaria, parliamentary)

(iv) Economy Relation: For the creation of economy relation, following pattern is created: relation_name (LOCATION, JJ). Afterwards, example sentence is chosen as follows: "The Netherlands has a market-based mixed economy, ranking 17th of 177 countries according to the Index of Economic Freedom." Therefore, the result of input had been:

economy(Netherlands, market-based)

(v) Geo Relation: For the creation of geo relation, following pattern is created: relation_name (LOCATION, CD). Afterwards, an example sentence is chosen as follows: "Egypt lies primarily between latitudes 22° and 32°N, and longitudes 25° and 35°E." Therefore, the result of input had been as follows:

geo(Egypt, 22°)
geo(Egypt, 32°N)
geo(Egypt, 25°)
geo(Egypt, 35°E)

(v) Religion Relation: For the creation of religion relation, following pattern is created: relation_name (LOCATION, NNS). Afterwards, an example sentence is chosen as follows: "Christianity is currently the largest religion in the Netherlands, accounting for about one-third of the population.". Therefore, the result of input had been as follows:

religion(Netherlands, Christianity is)

(vii) Member Relation: For the creation of member relation, following pattern is created: relation_name (LOCATION, ORGANIZATION). Afterwards, the example sentence is chosen as follows: "Germany is a member of UN, NATO, the G8, the G20 and the OECD." The result of that input had been:

member(Germany,UN)
member(Germany,NATO)
member(Germany,OECD./O)

(viii) Climate Relation: For the creation of climate relation, following pattern is created: relation_name (LOCATION, JJ). Afterwards, an example sentence is chosen as follows: " Ukraine has a mostly temperate continental climate, although the southern coast has a humid subtropical climate." Therefore, the result of input had been:

climate(Ukraine, mostly temperate continental)

(ix) State Relation: For the creation of state relation, following pattern is created: relation_name (LOCATION, JJ). Afterwards, an example sentence is chosen as follows: “France is a sovereign state comprising territory in western Europe and several overseas regions and territories.” Therefore, the result of input had been:

state(France, sovereign)

(x) Location Relation: For the creation of location relation, following pattern is created: relation_name (LOCATION, LOCATION). Afterwards, example sentence is chosen as follows: “ Italy is a unitary parliamentary republic in Europe.” Following that, the result of that input had been:

location(Italy,Europe)

(xi) Border Relation: For the creation of border relation, following pattern is created: relation_name (LOCATION, LOCATION). Afterwards, the example sentence is chosen as follows: “Turkey is bordered by eight countries: Syria and Iraq to the south; Iran, Armenia, and the Azerbaijani exclave of Nakhchivan to the east; Georgia to the northeast; Bulgaria to the northwest; and Greece to the west.” The result of that input had been:

border(Turkey,Syria)
border(Turkey,Iraq)
border(Turkey,Iran)
border(Turkey,Armenia)
border(Turkey,Georgia)
border(Turkey,Bulgaria)
border(Turkey,Greece)

(xii) Author Relation: For the creation of author relation, following pattern is created: relation_name(PERSON). Afterwards, an example sentence is chosen as follows: “Well-known German authors include Johann Wolfgang

von Goethe, Friedrich Schiller, Gotthold Ephraim Lessing and Theodor Fontane.". Therefore, the result of input had been as follows:

author(Johann)
author(Wolfgang)
author(Goethe)
author(Friedrich)
author(Schiller)
author(Gotthold)
author(Ephraim)
author(Lessing)
author(Theodor)
author(Fontane)

(xiii) Composer Relation: For the creation of composer relation, following pattern is created: relation_name (PERSON). Afterwards, an example sentence is chosen as follows: "Dieterich Buxtehude composed oratorios for organ, which influenced the later work of Johann Sebastian Bach and Georg Friedrich Händel; these men were influential composers of the Baroque period". Therefore, the result of input had been as follows:

composer(Dieterich)
composer(Buxtehude)
composer(Johann)
composer(Sebastian)
composer(Bach)
composer(Georg)
composer(Friedrich)

(xiv) Artist Relation: For the creation of artist relation, following pattern is created: relation_name(PERSON). Afterwards, an example sentence is chosen as follows: "French artists developed the rococo style in the 18th century, as a more intimate imitation of old baroque style, the works of the court-endorsed artists Antoine Watteau, François Boucher and Jean-Honoré

Fragonard being the most representative in the country.". Therefore, the result of input had been as follows:

artist(Antoine)
artist(Watteau)
artist(François)
artist(Boucher)
artist(Jean-Honoré)
artist(Fragonard)

(xv) Brand Relation: For the creation of brand relation, following pattern is created: relation_name(PERSON). Afterwards, an example sentence is chosen as follows: "The new car market is dominated by domestic brands such as Renault (27% of cars sold in France in 2003), Peugeot (20.1%) and Citroën (13.5%)." Therefore, the result of input had been as follows:

brand(Renault)
brand(Peugeot)
brand(Citroën)

(xvi) Physicist Relation: For the creation of physicist relation, following pattern is created: relation_name (PERSON). Afterwards, an example sentence is chosen as follows: "Notable German physicists before the 20th century include Hermann von Helmholtz, Joseph von Fraunhofer and Gabriel Daniel Fahrenheit, among others." Therefore, the result of input had been as follows:

physicist(Hermann)
physicist(Helmholtz)
physicist(Joseph)
physicist(Fraunhofer)
physicist(Gabriel)
physicist(Daniel)

(xvii) Mathematician Relation: For the creation of mathematician relation, following pattern is created: relation_name (PERSON). Afterwards, an example sentence is chosen as follows: "Galileo Galilei (1564–1642), a physicist, mathematician and astronomer, played a major role in the Scientific Revolution." Therefore, the result of input had been as follows:

mathematician(Galileo)

mathematician(Galilei)

The following table summarizes the relations with the grammatical patterns that are used with their meanings:

Table 3.6. Grammatical Patterns

Relation	**Grammatical Pattern**	**Meaning**
area	NP < CD	NP immediately dominates CD
population	NP < CD	NP immediately dominates CD
location	PP < NP	PP immediately dominates NP
economy	NP < JJ & <<NN	NP immediately dominates JJ and NN dominates them.
republic	NP < JJ & <<NN	NP immediately dominates JJ and NN dominates them.
state	NP < JJ & <<NN	NP immediately dominates JJ and NN dominates them.
religion	NP < NP	NP immediately dominates NP
climate	NP < JJ & <<NN	NP immediately dominates JJ and NN dominates them.
geo	CD	CD is taken from the sentence.

Creating relations had been consisted of using Named Entity Recognizer in addition to parsing sentences with Stanford NLP Parser. Table 3.7 shows usage of Named Entity Recognition (NER) in relations as well as their abstract representation by combining it with grammatical patterns.

Table 3.7. Abstract Representation of Relations and Named Entity Recognition

Relation	**NER**	**Abstract Representation**	**Example Relation**
border	LOCATION	border(LOCATION, pattern)	border(Turkey,Bulgaria)
area	LOCATION	area (LOCATION, pattern)	area(Italy,301,338)
member	ORGANIZATION	member(ORGANIZATION, pattern)	member(Germany,UN)

Table 3.7. Abstract Representation of Relations and Named Entity Recognition (cont.)

population	LOCATION	populatıon(LOCATION, pattern)	population(Bulgaria, 7,364,570)
location	LOCATION	location(LOCATION, LOCATION)	location(Greece,Europe)
economy	LOCATION	economy(LOCATION, pattern)	economy(Netherlands, market-based)
republic	LOCATION	republic(LOCATION, pattern)	republic(Bulgaria,parliamentary)
state	LOCATION	state(LOCATION, pattern)	state(France, sovereign)
religion	LOCATION	religion(LOCATION, pattern)	religion(Netherlands, Christianity is)
climate	LOCATION	climate(LOCATION, pattern)	climate(Denmark, temperate)
geo	LOCATION	geo(LOCATION, pattern)	geo(Egypt, 32°N)
author	PERSON	author(PERSON)	author(Goethe)
composer	PERSON	composer(PERSON)	composer(Bach)
artist	PERSON	artist(PERSON)	artist(Boucher)
physicist	PERSON	physicist(PERSON)	physicist(Helmholtz)
mathematician	PERSON	mathematician(PERSON)	mathematician(Galilei)
brand	ORGANIZATION	brand(ORGANIZATION)	brand(Peugeot)

When Table 3.6 and 3.7 are analyzed together, meaning of the whole relation could be inferred. Person, location and organization names had been an important part of generating relations.

4. EXPERIMENTS AND EVALUATION

In order to evaluate our relations, we tested each relation with ten sentences for each relation. Sentences containing keywords related to relation name are chosen in the first step. Afterwards, success of the system is measured based on the number of sentences and their outputs. Error analysis is done as shown on Table 4.1:

Table 4.1. Results of Relations

Relation Name	**A**	**B**	**C**	**D**	**E**	**F**	**Correctness-1**	**Correctness-2**
Area	6	1	0	1	2	10	0.60	0.66
Population	4	0	0	2	4	10	0.40	0.55
Republic	10	0	0	0	0	10	1.00	1.00
Economy	3	5	1	1	0	10	0.30	0.38
Geo	10	0	0	0	0	10	1.00	1.00
Religion	2	0	0	2	6	10	0.20	0.25
Member	9	0	0	1	0	10	0.90	1.00
Climate	4	0	1	0	5	10	0.40	0.44
State	8	0	0	2	0	10	0.80	1.00
Location	7	0	0	1	2	10	0.70	0.78
Border	3	3	4	0	0	10	0.30	0.50
Author	10	0	0	0	0	10	1.00	1.00
Composer	10	0	0	0	0	10	1.00	1.00
Artist	10	0	0	0	0	10	1.00	1.00
Brand	7	0	0	0	0	7	1.00	1.00
Mathematician	3	2	0	0	0	5	0.60	0.60
Physicist	6	1	0	0	0	7	0.86	0.86

The explanations of the column names in the table are as follows:

- A: The result is correct as depicted in Table 3.7.
- B: The result is partly correct. One of the arguments in the representation is not the correct answer, but the correct answer can be inferred.
- C: The result is incorrect. The error in the result is due to incorrect named entity output by the Stanford dependency parser.
- D: The result is incorrect. The error in the result is due to incorrect parse output by

the Stanford dependency parser.

- E: The result is incorrect. The grammatical pattern (Table 3.6) is not applicable for the sentence.
- F: Total number of sentences for the relation (i.e. A+B+C+D+E).

Correctness-1: A / F

Correctness-2: A / (F-C-D)

The first correctness measure (Correctness-1) shows the accuracy when all types of errors (B,C,D,E) are taken into account. However, the errors denoted by the columns C and D originate from the incorrect outputs of the Stanford tools, on which the approach in this paper is based. Therefore, we give an additional correctness measure (Correctness-2), in which the sentences parsed incorrectly are excluded. If we take 0.50 as a threshold for the success rate, we see that 12 out of 17 relations with respect to Correctness-1 (71%) and 14 out of 17 relations with respect to Correctness-2 (82%) are successful. Test results in detail could be find in Appendix part of this thesis. Below we give a brief analysis for each relation type.

For the area relation, the problem related to the parser output is caused by the representation of area information as adjective (JJ) rather than cardinal number (CD). An example sentence is the following: “Italy covers an area of 301,338 km2 (116,347 sq mi) and has a largely temperate seasonal climate; due to its shape, it is often referred to in Italy as lo Stivale (the Boot).”.

For the population relation, an example with grammatical pattern problem is: “Egypt is the most populated country in the Middle East, and the third most populous on the African continent, with about 88 million inhabitants as of 2015.” The output of this relation is given as “population (Egypt, 2015)”.

For the republic and geo relations, the identified patterns were shown to be suitable for the sentence structures and thus all results were correct.

For the economy relation, half of the results were partly correct, which is related to both patterns and sentence structures. An example sentence of economy relation is as follows: “Italy has a capitalist mixed economy, ranking as the third-largest in the Eurozone and the eighth-largest in the world.”. This sentence resulted in the output “economy (Italy,

third-largest)".

For the religion and climate relations, it was quite challenging to generate a general rule since the sentences take highly different forms. Thus, a significant number of pattern errors were observed. For the member and state relations, errors were due to problems in the Stanford parser, which can be ignored for the evaluation. For the location relation, a small number of errors were caused by the pattern rule.

For the border relation, named entity problems occurred due to the mixed structure of the sentences. An example of this is the relation "border (Ukraine, bordered by Russia to)", extracted from the sentence "Ukraine bordered by Russia to the east and northeast, Belarus to the northwest, Poland and Slovakia to the west, Hungary, Romania, and Moldova to the southwest, and the Black Sea and Sea of Azov to the south and southeast, respectively.".

For the relations of author, composer, artist, brand, mathematician and physicist, no errors were detected.

5. CONCLUSION

In this work, a method for converting a natural language text into semantic relations was proposed. By using Stanford dependency parser and named entity recognizer, 17 different relations were identified by using sentences from Wikipedia documents in geography domain. The success rates of the relations ranged between 20% and 100%. It was shown for 71% of the relations more than half of the outputs were correct. When the errors related to the parser and named entity recognizer are ignored, the success rates ranged between 25% and 100%, and the percentage of relations above the 0.50 threshold raised to 82%.

Based on the results, we can conclude that text written in natural language in a particular domain can be converted into relations, containing meaningful information derived from the text. We identified the importance of the relations by analyzing sentences manually on Wikipedia pages and chose the sentences accordingly. For future work, relations on different domains can be considered. Moreover, approaches based on syntactical trees and semantic information can be developed.

REFERENCES

Bajwa I.S. , Lee M. , Bordbar B., "Translating Natural Language Constraints to OCL" Journal of King Saud University - Computer and Information Sciences, Vol. 24 (2), pp. 117–128, 2012.

Cheng J., Ma Z.M., Yhan L., Zhang F.,"Knowledge Representation and Reasoning of XML with Ontology, SAC '11 Proceedings of the 2011 ACM Symposium on Applied Computing. pp. 1705-1710, 2011.

Dinesh N., Joshi A., Lee I., Extracting Formal Specifications from Natural Language Regulatory Documents. Retrieved from http://www.aclweb.org/anthology/W06-3902, accessed at April 2016.

Drechsler R., Harris C.B., Harris I.G., Abdessaied N., Soeken M., "Automating the Translation of Assertions Using Natural Language Processing Techniques". IEEE Forum on Specification and Design Languages. Vol. 978,pp. 1-8, 2014.

Drechsler R., Soeken M., Wille R. "Formal Specification Level: Towards Verification-Driven Design Based on NLP", IEEE Forum on Specification and Design Languages (FDL), pp. 53 – 58, 2012.

Elenius D., Ghosh S., Li W., Lincoln P., Shankar N., Steiner W., "ARSENAL-Automatically Extracting Requirements Specifications from Natural Language", 2014, Retrieved from: http://arxiv.org/abs/1403.3142

Harris C.B., Harris I.G., "Generating Formal Hardware Verification Properties from Natural Language Documentation", *IEEE Conference on Semantic Computing*, pp. 49-56, 2015.

Harris I. G., Capturing assertions from natural language descriptions. *Natural Language Analysis in Software Engineering (NaturaLiSE)*, pp. 17-24, 2013.

Hermjakob U., Knight K., Marcu D., May J., Pust M., "Parsing English into Abstract Meaning Representation Using Syntax-Based Machine Translation." *Proceedings of EMNLP*, pp. 1143-1154, 2015.

Holt A., "Formal Verification with Natural Language Specifications: spectrum guidelines, experiments, lessons so far"., *South African Computer Journal* , No. 24, pp. 253–257, 1999.

Klein D., Manning C.D., " Accurate Unlexicalized Parsing". *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, pp. 423-430, 2003.

Lee B., Bryant B.R., "Contextual Knowledge Representation for Requirements Documents in Natural Language*", FLAIRS-02 Proceedings*, 2002.

Lee B., Bryan B.R., “Automated Conversion from Requirements Documentation to an Object-Oriented Formal Specification Language” , *SAC'02 Proceedings of the 2002 ACM Symposium on Applied Computing*, pp. 932-936, 2002.

Mencl V. , “Deriving Behaviour Specifications from Textual Use Cases”, 2004, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.11.312, accessed at June 2016.

Stuckenschmidt H. , Schuster G. , Vögele T. , Visser U., “Ontologies for geographic information processing, Computers & Geosciences”, Vol.28(1). pp. 103-117, 2002. http://www.sciencedirect.com/science/article/pii/S009830040100019X, accessed at June 2016.

Wang W.M., Cheung C. F. , Lee W.B., Kwok S.K., “Mining knowledge from natural language texts using fuzzy associated concept mapping”, *ACM Journal of Information Processing and Management: and International Journal*, vol.44(5), pp. 1707-1719, 2008.

Stanford Dependencies, http://nlp.stanford.edu/software/stanford-dependencies.shtml, accessed at April 2016.

Stanford Lexical Parser, http://nlp.stanford.edu/software/lex-parser.shtml, accessed at April 2016.

Stanford Named Entity Recognizer, http://nlp.stanford.edu/ner/, accessed at May 2016.

TregexPattern, http://nlp.stanford.edu/nlp/javadoc/javanlp/edu/stanford/nlp/trees/tregex/TregexPattern.html, accessed at April 2016.

APPENDIX A : RESULTS

In this Appendix, details of evaluation as well as selected sentences are listed. In the first line after relation names, chosen sentences are shown. Afterwards, outputs of the programs are listed. Following that, error and error types are listed due to Table 4.1.

A.1. Area Relation

- Germany includes 16 constituent states and covers an area of 357,021 square kilometres (137,847 sq mi) with a largely temperate seasonal climate.
  (i) Output:
  area(Germany, 16)
  (ii) Error:
  Result is partly true, as Germany is inside of the relation, whereas number is incorrect.
  Error Type: B

- Italy covers an area of 301,338 km2 (116,347 sq mi) and has a largely temperate seasonal climate; due to its shape, it is often referred to in Italy as lo Stivale (the Boot).
  (i) Error:
  Number is listed as JJ rather than CD, where the number is 301,338.
  Error Type: D

- At 17,125,200 square kilometres (6,612,100 sq mi), Russia is the largest country in the world, covering more than one-eighth of the Earth's inhabited land area.
  (i) Output:
  area(Russia, 17,125,200)
  (ii) Error:
  No Error

- With an area of 505,990 km^2 (195,360 sq mi), Spain is the second largest country in

Western Europe and the European Union, and the fourth largest country in Europe.

(i) Output:

area(Spain, 505,990 km2 195,360 sq mi,)

(ii) Error:

No Error

- Turkey's area, including lakes, occupies 783,562 square kilometres (302,535 square miles),of which 755,688 square kilometres (291,773 square miles) are in Southwest Asia and 23,764 square kilometres (9,175 square miles) in Europe.

(i) Output:

area(Turkey's/O, 783,562)

(ii) Error:

No Error

- With a territory of 110,994 square kilometres (42,855 sq mi), Bulgaria is Europe's 16th-largest country.

(i) Output:

area(Bulgaria, 110,994)

(ii) Error:

No Error

- France spans 643,801 square kilometres (248,573 sq mi) and has a total population of 66.6 million.

(i) Output:

area(France, 643,801)

(ii) Error:

No Error

- The total area of Poland is 312,679 square kilometres (120,726 sq mi), making it the 69th largest country in the world and the 9th largest in Europe.

(i) Output:

area(Poland is/O,NNP Poland is)

(ii) Error:

Pattern is not applicable for the sentence, giving the wrong output.

Error Type: E

- Including Crimea, Ukraine has an area of 603,628 km^2 (233,062 sq mi), making it the largest country entirely within Europe and the 46th largest country in the world, and a population of about 44.5 million, making it the 32nd most populous country in the world.

(i) Error:

Pattern is not applicable for the sentence.

Error Type: E

- With an area of 93,800 square miles (243,000 km^2), the UK is the 80th-largest sovereign state in the world and the 11th-largest in Europe.

(i) Output:

area(UK, 93,800)

(ii) Error:

No Error

A.2. Population Relation

- With about 81.5 million inhabitants, Germany is the most populous member state in the European Union.

(i) Output:

population(Germany,RB about)

(ii) Error:

Pattern is not applicable for the sentence.

Error Type: E

- With 61 million inhabitants, Italy is the 4th most populous EU member state.

(i) Error:

Pattern is not applicable for the sentence.

Error Type: E

- Russia is the world's ninth most populous country with over 144 million people at the end of 2015.
  (i) Output:
  population(Russia, 2015.)
  (ii) Error:
  Result of the pattern is wrong.
  Error Type: E

- The Netherlands had an estimated population of 16,785,403 on 30 April 2013.
  (i) Output:
  population(Netherlands, 16,785,403)
  (ii) Error:
  No Error.

- The population of Bulgaria is 7,364,570 people according to the 2011 national census.
  (i) Output:
  population(Bulgaria, 7,364,570)
  (ii) Error:
  No Error.

- France spans 643,801 square kilometres (248,573 sq mi) and has a total population of 66.6 million.
  (i) Output:
  population(France, 643,801)
  (ii) Error:
  No Error.

- With a population of over 38.5 million people, Poland is the 34th most populous country in the world, the 8th most populous country in Europe and the sixth most

populous member of the European Union, as well as the most populous post-communist member of the European Union.

(i) Error:

Parsing error.

Error Type: D

- With 19.94 million inhabitants, Romania is the seventh most populous member state of the European Union.

(i) Output:

population(Romania, 19.94)

(ii) Error:

No Error.

- According to the *Address-Based Population Recording System* of Turkey, the country's population was 74.7 million people in 2011, nearly three-quarters of whom lived in towns and cities.

(i) Output:

population(Turkey,/O, 2011,)

(ii) Error:

Parsing Error.

Error Type: D

- Egypt is the most populated country in the Middle East, and the third most populous on the African continent, with about 88 million inhabitants as of 2015.

(i) Output:

population(Egypt, 2015.)

(ii) Error:

Pattern is not applicable for the sentence.

Error Type: E

A.3. Republic Relation

- Germany is a federal parliamentary republic in West-Central Europe.
  (i) Output:
  republic(Germany, federal)
  (ii) Error:
  No Error.

- Italy is a unitary parliamentary republic in Europe.
  (i) Output:
  republic(Italy, unitary)
  (ii) Error:
  No Error.

- Russia is a federal semi-presidential republic.
  (i) Output:
  republic(Russia, federal)
  (ii) Error:
  No Error.

- Turkey is a parliamentary republic in Eurasia, largely located in Western Asia, with the smaller portion of Eastern Thrace in Southeast Europe.
  (i) Output:
  republic(Turkey, parliamentary)
  (ii) Error:
  No Error.

- Greece is a unitary parliamentary republic.
  (i) Output:
  republic(Greece, a unitary parliamentary)
  (ii) Error:
  No Error.

- France is a unitary semi-presidential republic with the capital in Paris, the country's largest city and main cultural and commercial centre.

  (i) Output:

  republic(France, unitary)

  (ii) Error:

  No Error.

- Romania is a semi-presidential republic where executive functions are held by both government and the president.

  (i) Output:

  republic(Romania, semi-presidential)

  (ii) Error:

  No Error.

- Estonia is a democratic parliamentary republic divided into fifteen counties, with its largest city and capital being Tallinn.

  (i) Output:

  republic(Estonia, democratic)

  (ii) Error:

  No Error.

- Portugal has been a semi-presidential representative democratic republic since the ratification of the Constitution of 1976, with Lisbon, the nation's largest city, as its capital.

  (i) Output:

  republic(Portugal, semi-presidential)

  (ii) Error:

  No Error.

- Finland is a parliamentary republic with a central government based in the capital Helsinki, local governments in 317 municipalities, and an autonomous region, the Åland Islands.

(i) Output:

republic(Finland, parliamentary)

(ii) Error:

No Error.

A.4. Economy Relation

- Italy has a capitalist mixed economy, ranking as the third-largest in the Eurozone and the eighth-largest in the world.

(i) Output:

economy(Italy, third-largest)

(ii) Error:

Result is partly true.

Error Type: B

- Russia has a developed, high-income market economy with enormous natural resources, particularly oil and natural gas.

(i) Error:

Problem due to Parser occurs.

Error Type: D

- Spain's capitalist mixed economy is the 16th largest worldwide and the 5th largest in the European Union, as well as the Eurozone's 4th largest.

(i) Error:

Problem due to Named Entity Recognizer occurs.

Error Type: C

- France has a mixed economy that combines extensive private enterprise with substantial state enterprise and government intervention.

(i) Output:

economy(France, extensive)

(ii) Error:

Result is partly true.

Error Type: B

- Poland's high-income economy is considered to be one of the healthiest of the post-Communist countries and is one of the fastest growing within the EU.

(i) Output:

economy(Poland's/O, Poland's)

(ii) Error:

Result is partly true.

Error Type: B

- Germany has a social market economy with a highly skilled labour force, a large capital stock, a low level of corruption, and a high level of innovation.

(i) Output:

economy(Germany, social)

(ii) Error:

No Error.

- The Netherlands has a market-based mixed economy, ranking 17th of 177 countries according to the Index of Economic Freedom.

(i) Output:

economy(Netherlands, market-based)

(ii) Error:

No Error.

- Ukraine is regarded as a developing economy with high potential for future success, though such a development is thought likely only with new all-encompassing economic and legal reforms.

(i) Output:

economy(Ukraine, high)

(ii) Error:

Result is partly true.

Error Type: B

- Austria is the 12th richest country in the world in terms of GDP (Gross domestic product) per capita, has a well-developed social market economy, and a high standard of living.

(i) Output:

economy(Austria, 12^{th} richest)

(ii) Error:

Result is partly true.

Error Type: B

- Belgium is a developed country, with an advanced high-income economy and is categorized as very high in the Human Development Index.

(i) Output:

economy(Belgium, advanced)

(ii) Error:

No Error.

A.5. Geo Relation

- It lies mostly between latitudes 47° and 55° N and longitudes 5° and 16° E.

(The word “it” is changed with Germany, in order to get relation.)

(i) Output:

geo(Germany, 47°)

geo(Germany, 55°)

geo(Germany, 5°)

geo(Germany, 16°)

(ii) Error:

No Error.

- Metropolitan France is situated mostly between latitudes 41° and 51° N, and longitudes 6° W and 10° E, on the western edge of Europe, and thus lies within the northern temperate zone.

  (i) Output:

  geo(France, 41°)
  geo(France, 51°)
  geo(France, 6°)
  geo(France, 10°)

  (ii) Error:

  No Error.

- Spain lies between latitudes 26° and 44° N, and longitudes 19° W and 5° E.

  (i) Output:

  geo(Spain, 26°)
  geo(Spain, 44°)
  geo(Spain, 19°)
  geo(Spain, 5°)

  (ii) Error:

  No Error.

- It lies between latitudes 35° and 43° N, and longitudes 25° and 45° E.

  (The word "it" is changed with Turkey, in order to get the relation.)

  (i) Output:

  geo(Turkey, 35°)
  geo(Turkey, 43°)
  geo(Turkey, 25°)
  geo(Turkey, 45°)

  (ii) Error:

  No Error.

- It lies between latitudes 41° and 82° N, and longitudes 19° E and 169° W.

  (The word "it" is changed with Russia, in order to get the relation.)

(i) Output:

geo(Russia, 41°)
geo(Russia, 82°)
geo(Russia, 19°)
geo(Russia, 169°)

(ii) Error:

No Error.

- Norway lies between latitudes 57° and 81° N, and longitudes 4° and 32° E.

(i) Output:

geo(Norway, 57°)
geo(Norway, 81°)
geo(Norway, 4°)
geo(Norway, 32°)

(ii) Error:

No Error.

- Poland's territory extends across several geographical regions, between latitudes 49° and 55° N, and longitudes 14° and 25° E.

(i) Output:

geo(Poland's/O, 49°)
geo(Poland's/O, 55°)
geo(Poland's/O, 14°)
geo(Poland's/O, 25°)

(ii) Error:

No Error.

- It lies between latitudes 44° and 53° N, and longitudes 22° and 41° E.

(The word “it” is changed with Ukraine, in order to get the relation.)

(i) Output:

geo(Ukraine, 44°)
geo(Ukraine, 53°)

geo(Ukraine, 22°)

geo(Ukraine, 41°)

(ii) Error:

No Error.

- Austria lies between latitudes 46° and 49° N, and longitudes 9° and 18° E.

(i) Output:

geo(Poland's/O, 49°)

geo(Poland's/O, 55°)

geo(Poland's/O, 14°)

geo(Poland's/O, 25°)

(ii) Error:

No Error.

- The European area of the Netherlands lies between latitudes 50° and 54° N, and longitudes 3° and 8° E.

(i) Output:

geo(Netherlands, 50°)

geo(Netherlands, 54°)

geo(Netherlands, 3°)

geo(Netherlands, 8°)

(ii) Error:

No Error.

A.6. Religion Relation

- According to the 2011 German Census, Christianity is the largest religion in Germany, claiming 66.8% of the total population.

(i) Output:

religion(Germany, German)

(ii) Error:

Problem due to grammatical pattern occurs.

Error Type:E

- Islam is the dominant religion of Turkey with 99.8 percent of the population being registered as Muslim.

  (i) Error:

  Problem due to Parsing occurs, as Islam is tagged as (VP (VBZ Islam is).

Error Type:D

- Christianity is currently the largest religion in the Netherlands, accounting for about one-third of the population.

  (i) Output:

  religion(Netherlands, Christianity is)

  (ii) Error:

  No Error.

- Catholicism has been the predominant religion in France for more than a millennium, though it is not as actively practiced today as it was.

  (i) Output:

  religion(France, Catholicism has)

  (ii) Error:

  No Error.

- Roman Catholicism is, by far, the largest religion in Italy, although Catholicism is no longer officially the state religion.

  (i) Error:

  Problem due to grammatical pattern occurs.

Error Type:E

- Several important cultural changes occurred during this time. Under Swedish and largely German rule, western Latvia adopted Lutheranism as its main religion.

(i) Output:

religion(Latvia, changes)

(ii) Error:

Problem due to Parser occurs.

Error Type:D

- Roman Catholicism has long been the main religion of Spain, and although it no longer has official status by law, in all public schools in Spain students have to choose either a religion or ethics class, and Catholicism is the only religion officially taught.

(i) Output:

religion(Spain,)

(ii) Error:

Problem due to grammatical pattern occurs.

Error Type:E

- In August 2012, ARENA estimated that about 46.8% of Russians are Christians (including Orthodox, Catholic, Protestant, and non-denominational), while 25% believed in God but without any religion.

(i) Error:

Problem due to grammatical pattern occurs.

Error Type:E

- According to new polls about Religiosity in the European Union in 2012 by Eurobarometer found that Christianity is the largest religion in Estonia accounting 28.06% of Estonians.

(i) Output:

religion(Estonia, polls)

(ii) Error:

Problem due to grammatical pattern occurs.

*Error Type:*E

- The 2011 census reported that Hinduism (79.8% of the population) is the largest

religion in India, followed by Islam (14.23%). Other religions or none (5.97% of the population) include Christianity (2.30%), Sikhism (1.72%), Buddhism (0.70%), Jainism, Judaism, Zoroastrianism, and the Bahá'í Faith.

(i) Output:

religion(India, religions)

(ii) Error:

Problem due to grammatical pattern occurs.

Error Type: E

A.7. Member Relation

- Germany is a member of UN, NATO, the G8, the G20 and the OECD.

(i) Output:

member(Germany,UN)
member(Germany,NATO)
member(Germany,OECD./O)

(ii) Error:

No Error.

- Italy is a founding and leading member of the European Union and the member of numerous international institutions including the UN, NATO, the OECD, the OSCE, the WTO , the G7/G8 , G20, the Union for the Mediterranean, the Council of Europe Uniting for Consensus and many more.

(i) Error:

Problem due to Parser occurs.

Error Type: D

- Spain is a member of UN, EU,CoE, OEI , NATO, OECD, WTO and many other international organisations.

(i) Output:

member(Spain,UN)
member(Spain,EU CoE/O)

member(Spain,NATO)

member(Spain,OECD)

member(Spain,WTO)

(ii) Error:

No Error.

- Turkey is a member of the UN, NATO, OECD, OSCE, OIC and the G-20.

(i) Output:

member(Turkey,the/O UN NATO OECD/O OSCE/O OIC/O and/O)

(ii) Error:

No Error.

- The Netherlands is a founding member of the EU, Eurozone, G-10, NATO, OECD and WTO, and a part of the trilateral Benelux Union.

(i) Output:

member(Netherlands,the/O EU Eurozone/O G-10/O NATO OECD and/O WTO)

No Error.

- Greece is also a member of numerous other international institutions, including the Council of Europe, NATO, OECD, OIF, OSCE and the WTO.

(i) Output:

member(Greece, OECD OIF/O OSCE and/O)

member(Greece,the/O WTO./O)

(ii) Error:

No Error.

- France is a member of the Group of 7, NATO, OECD, WTO and La Francophonie.

(i) Output:

member(France,7/O NATO OECD/O)

member(France,WTO)

(ii) Error:

No Error.

- Russia is a great power and a permanent member of the United Nations Security Council, a member of the G20, the Council of Europe, the Asia-Pacific Economic Cooperation (APEC), the Shanghai Cooperation Organisation (SCO), the Organization for Security and Co-operation in Europe (OSCE), and the World Trade Organization (WTO), as well as being the leading member of the Commonwealth of Independent States (CIS), the Collective Security Treaty Organization (CSTO) and one of the 5 members of the Eurasian Economic Union (EEU), along with Armenia, Belarus, Kazakhstan and Kyrgyzstan.

(i) Output:

member(Russia,-LRB-/OAPEC-RRB-/O)
member(Russia,-LRB-/OSCO-RRB-/O)
member(Russia,-LRB-/OOSCE-RRB-/O)
member(Russia,-LRB-/OWTO-RRB-/O)
member(Russia,-LRB-/OCIS-RRB-/O)

(ii) Error:

No Error.

- The Netherlands is a founding member of the EU, Eurozone, G-10, NATO, OECD and WTO, and a part of the trilateral Benelux Union.

(i) Output:

member(Netherlands,EU)
member(Netherlands,NATO)
member(Netherlands,OECD)
member(Netherlands,WTO)

(ii) Error:

No Error.

- Portugal is a member of numerous international organizations, including the United

Nations, the European Union, the eurozone, OECD, NATO and the Community of Portuguese Language Countries.

(i) Output:

member(Portugal,OECD)

member(Portugal,NATO)

(ii) Error:

No Error.

A.8. Climate Relation

- Most of Germany has a temperate seasonal climate dominated by humid westerly winds.

(i) Output:

climate(Germany, temperate)

(ii) Error:

No Error.

- The enormous size of Russia and the remoteness of many areas from the sea result in the dominance of the humid continental climate, which is prevalent in all parts of the country except for the tundra and the extreme southeast.

(i) Output:

climate(Russia, enormous)

(ii) Error:

Problem with pattern occurs.

Error Type: E

- Ukraine has a mostly temperate continental climate, although the southern coast has a humid subtropical climate.

(i) Output:

climate(Ukraine, mostly temperate continental)

(ii) Error:

No Error.

- Thanks to the great longitudinal extension of the peninsula and the mostly mountainous internal conformation, the climate of Italy is highly diverse.

  (i) Output:

  climate(Italy, great)

  (ii) Error:

  Problem with pattern occurs.

  Error Type: E

- The climate of Greece is primarily Mediterranean, featuring mild, wet winters and hot, dry summers.

  (i) Error:

  Problem with pattern occurs.

  Error Type: E

- The coastal areas of Turkey bordering the Aegean and Mediterranean Seas have a temperate Mediterranean climate, with hot, dry summers and mild to cool, wet winters.

  (i) Output:

  climate(Turkey, coastal)

  (ii) Error:

  Problem with pattern occurs.

  Error Type: E

- The enormous size of Russia and the remoteness of many areas from the sea result in the dominance of the humid continental climate, which is prevalent in all parts of the country except for the tundra and the extreme southeast.

  (i) Output:

  climate(Russia, enormous)

  (ii) Error:

  Problem with pattern occurs.

  Error Type: E

- The islands of the Caribbean Netherlands enjoy a tropical climate with warm weather all year round.

  (i) Output:

  climate(Caribbean, tropical)

  (ii) Error:

  Problem occurs due to the output generated by Stanford Named Entity Recognizer.

  Error Type: C

- Portugal is defined as a Mediterranean climate.

  (i) Error:

  Problem with pattern occurs.

  Error Type: E

- Denmark has a temperate climate, characterised by mild winters, with mean temperatures in January of 1.5 °C (34.7 °F), and cool summers, with a mean temperature in August of 17.2 °C (63.0 °F).

  (i) Output:

  climate(Denmark, temperate)

  (ii) Error:

  No Error.

A.9. State Relation

- Finland is a sovereign state in Europe.

  (i) Output:

  state(Finland, sovereign)

  (ii) Error:

  No Error.

- Turkey is a secular state with no official state religion; the Turkish Constitution

provides for freedom of religion and conscience.

(i) Output:

state(Turkey, official)

(ii) Error:

Problem occurs due to Parser error.

Error Type: D

- Russia is a sovereign state in northern Eurasia.

(i) Output:

state(Russia, sovereign)

(ii) Error:

No Error.

- Spain is a sovereign state largely located on the Iberian Peninsula in southwestern Europe, with archipelagos in the Atlantic Ocean and Mediterranean Sea, and several small territories on and near the north African coast.

(i) Output:

state(Spain, sovereign)

(ii) Error:

No Error.

- Romania is a secular state and has no state religion.

(i) Output:

state(Romania, no state)

(ii) Error:

Problem occurs due to Parser error.

Error Type: D

- The United Kingdom of Great Britain and Northern Ireland, is a sovereign state in Europe.

(i) Output:

state(Britain, sovereign)

(ii) Error:

No Error.

- Belgium is a sovereign state in Western Europe.

(i) Output:

state(Belgium, sovereign)

(ii) Error:

No Error.

- France is a sovereign state comprising territory in western Europe and several overseas regions and territories.

(i) Output:

state(France, sovereign)

(ii) Error:

No Error.

- The Netherlands is described as a consociational state.

(i) Output:

state(Netherlands, consociational)

(ii) Error:

No Error.

- Norway, a unitary state, is divided into nineteen first-level administrative counties (fylke).

(i) Output:

state(Norway, unitary)

(ii) Error:

No Error.

A.10. Location Relation

- Austria is a federal republic and a landlocked country of over 8.66 million

people in Central Europe.

(i) Error:

Problem occurs due to Parser error.

Error Type: D

- Japan is an island country in East Asia.

(i) Output:

location(Japan,Asia)

(ii) Error:

No Error.

- Russia is a sovereign state in northern Eurasia.

(i) Output:

location(Russia,Eurasia)

(ii) Error:

No Error.

- Germany is a federal parliamentary republic in west-central Europe.

(i) Output:

location(Germany,Europe)

(ii) Error:

No Error.

- Italy is a unitary parliamentary republic in Europe.

(i) Output:

location(Italy,Europe)

(ii) Error:

No Error.

- Spain is a sovereign state largely located on the Iberian Peninsula in southwestern Europe, with archipelagos in the Atlantic Ocean and Mediterranean Sea, and several small territories on and near the north African coast.

(i) Error:

Problem occurs due to grammatical pattern.

Error Type: E

- Turkey is a parliamentary republic in Eurasia, largely located in Western Asia, with the smaller portion of Eastern Thrace in Southeast Europe.

(i) Output:

location(Turkey,Eurasia,/O)

(ii) Error:

No Error.

- Netherlands is a small, densely populated country located in Western Europe with three island territories in the Caribbean.

(i) Output:

location(Netherlands,Europe)

(ii) Error:

No Error.

- Greece is a country located in southeastern Europe.

(i) Output:

location(Greece,Europe)

(ii) Error:

No Error.

- The total area of Poland is 312,679 square kilometres (120,726 sq mi), making it the 69th largest country in the world and the 9th largest in Europe.

(i) Error:

Problem occurs due to grammatical pattern.

Error Type: E

A.11. Border Relation

- From northwest to southeast, Russia shares land borders with Norway, Finland, Estonia, Latvia, Lithuania and Poland (both with Kaliningrad Oblast), Belarus,Ukraine, Georgia, Azerbaijan, Kazakhstan, China, Mongolia, and North Korea.

  (i) Output:

border(Russia shares/O,borders/O with/O Norway)
border(Russia shares/O, Finland)
border(Russia shares/O, Estonia)
border(Russia shares/O, Latvia)
border(Russia shares/O, Lithuania and/O Poland-LRB-/Oboth/O)
border(Russia shares/O,with/O Kaliningrad)
border(Russia shares/O, Belarus)
border(Russia shares/O,Ukraine)
border(Russia shares/O, Georgia)
border(Russia shares/O, Azerbaijan)
border(Russia shares/O, Kazakhstan)
border(Russia shares/O, China)
border(Russia shares/O, Mongolia)
border(Russia shares/O,Korea./O)
border(Russia shares/O,Russia shares/O)
border(Russia shares/O,borders/O with/O Norway)
border(Russia shares/O, Finland)
border(Russia shares/O, Estonia)
border(Russia shares/O, Latvia)
border(Russia shares/O, Lithuania and/O Poland -LRB-/Oboth/O)
border(Russia shares/O,with/O Kaliningrad)
border(Russia shares/O, Belarus)
border(Russia shares/O,Ukraine)
border(Russia shares/O, Georgia)

border(Russia shares/O, Azerbaijan)
border(Russia shares/O, Kazakhstan)
border(Russia shares/O, China)
border(Russia shares/O, Mongolia)
border(Russia shares/O,Korea./O)

(ii) Error:

Result is incorrect due to incorrect output of Stanford Named Entity Recognizer.

Error Type: C

- Austria is bordered by the Czech Republic and Germany to the north, Hungary and Slovakia to the east, Slovenia and Italy to the south, and Switzerland and Liechtenstein to the west.

(i) Output:

border(Austria,and/O Germany to/O)
border(Austria, Hungary and/O Slovakia to/O)
border(Austria, Slovenia and Italy/O to/O)
border(Austria,and/O Switzerland and/O Liechtenstein to/O)

(ii) Error:

Result is incorrect due to incorrect output of Stanford Named Entity Recognizer.

Error Type: C

- Ukraine bordered by Russia to the east and northeast, Belarus to the northwest, Poland and Slovakia to the west, Hungary, Romania, and Moldova to the southwest, and the Black Sea and Sea of Azov to the south and southeast, respectively.

(i) Output:

border(Ukraine,bordered/O by/O Russia to/O)
border(Ukraine, Belarus to/O)
border(Ukraine,Poland and/O Slovakia to/O)
border(Ukraine,Hungary)

border(Ukraine,Romania)

border(Ukraine,and/O Moldova to/O)

(ii) Error:

Result is incorrect due to incorrect output of Stanford Named Entity Recognizer.

Error Type: C

- Poland is a country in Central Europe, bordered by Germany to the west; the Czech Republic and Slovakia to the south; Ukraine and Belarus to the east; and the Baltic Sea,Kaliningrad Oblast (a Russian exclave) and Lithuania to the north.

(i) Output:

border(Poland ,Europe)

border(Poland , bordered/O by/O Germany to/O)

border(Poland ,Slovakia to/O)

border(Poland ,south/O;/O Ukraine and/O Belarus to/O)

border(Poland ,Kaliningrad)

border(Poland ,and/O Lithuania to/O)

border(Poland ,Poland)

border(Poland ,Europe)

border(Poland ,bordered/O by/O Germany to/O)

border(Poland ,Slovakia to/O)

border(Poland ,south/O;/O Ukraine and/O Belarus to/O)

border(Poland ,Kaliningrad)

border(Poland,and/O Lithuania to/O)

- Located in the heart of the Mediterranean Sea, Italy shares open land borders with France, Switzerland, Austria, Slovenia, San Marino and Vatican City.

(i) Output:

border(Mediterranean,Italy)

border(Mediterranean,France)

border(Mediterranean,Switzerland)

border(Mediterranean,Austria)

border(Mediterranean,Slovenia)

border(Mediterranean,Mediterranean)

border(Mediterranean,Italy)

border(Mediterranean,France)

border(Mediterranean,Switzerland)

border(Mediterranean,Austria)

border(Mediterranean,Slovenia)

(ii) Error:

Result is partly correct.

Error Type: B

- Bounded by the Indian Ocean on the south, the Arabian Sea on the south-west, and the Bay of Bengal on the south-east, it shares land borders with Pakistan to the west; China, Nepal, and Bhutan to the north-east; and Myanmar (Burma) and Bangladesh to the east.

(The word "it" is replaced by "India" in order to make relation correct.)

(i) Output:

border(Bengal,India)

border(Bengal,Pakistan)

border(Bengal,China)

border(Bengal,Nepal)

border(Bengal,Bhutan)

border(Bengal,Myanmar)

border(Bengal,-LRB-/OBurma-RRB-/O)

border(Bengal,Bangladesh)

border(Bengal,Bengal)

border(Bengal,India)

border(Bengal,Pakistan)

border(Bengal,China)

border(Bengal,Nepal)

border(Bengal,Bhutan)

border(Bengal,Myanmar)

border(Bengal,-LRB-/OBurma-RRB-/O)

border(Bengal,Bangladesh)

(ii) Error:

Result is partly correct.

Error Type: B

- Situated on the southern tip of the Balkan peninsula, it shares land borders with Albania to the northwest, the Republic of Macedonia and Bulgaria to the north and Turkey to the northeast.

  (The word "it" is replaced by Greece for the relation.)

  (i) Output:

  border(Balkan,Greece)

  border(Balkan,Albania)

  border(Balkan,Macedonia)

  border(Balkan,Bulgaria)

  border(Balkan,Turkey)

  border(Balkan,Balkan)

  border(Balkan,Greece)

  border(Balkan,Albania)

  border(Balkan,Macedonia)

  border(Balkan,Bulgaria)

  border(Balkan,Turkey)

  (ii) Error:

  Result is partly correct.

  Error Type: B

- It is bordered to the northwest by Armenia, the *de facto* Nagorno-Karabakh Republic, and Azerbaijan; to the north by Kazakhstan and Russia across the Caspian Sea; to the northeast by Turkmenistan; to the east by Afghanistan and Pakistan; to the south by the Persian Gulf and the Gulf of Oman; and to the west by Turkey and Iraq.

  (The word "it" is replaced by "Iran" in order to make relation correct.)

(i) Output:

border(Iran,Armenia)
border(Iran,Nagorno-Karabakh)
border(Iran,Azerbaijan;/O)
border(Iran,Kazakhstan)
border(Iran,Russia)
border(Iran,Turkmenistan;/O)
border(Iran,Afghanistan)
border(Iran,Pakistan;/O)
border(Iran,Oman;/O)
border(Iran,Turkey)
border(Iran,Iraq./O)
border(Iran,Iran)
border(Iran,Armenia)
border(Iran,Nagorno-Karabakh)
border(Iran,Azerbaijan;/O)
border(Iran,Kazakhstan)
border(Iran,Russia)
border(Iran,Turkmenistan;/O)
border(Iran,Afghanistan)
border(Iran,Pakistan;/O)
border(Iran,Oman;/O)
border(Iran,Turkey)
border(Iran,Iraq./O)

(ii) Error:

No Error.

- Turkey is bordered by eight countries: Syria and Iraq to the south; Iran, Armenia, and the Azerbaijani exclave of Nakhchivan to the east; Georgia to the northeast; Bulgaria to the northwest; and Greece to the west.

(i) Output:

border(Turkey,Syria)

border(Turkey,Iraq)

border(Turkey,Iran)

border(Turkey,Armenia)

border(Turkey,Georgia)

border(Turkey,Bulgaria)

border(Turkey,Greece)

(ii) Error:

No Error.

- It is bordered to the north by the Gulf of Finland, to the west by the Baltic Sea, to the south by Latvia (343 km), and to the east by Lake Peipus and Russia (338.6 km).

  (The word “it” is replaced by the word “Estonia” to make relation correct.)

  (i) Output:

border(Estonia,Finland)

border(Estonia,Latvia)

border(Estonia,Russia)

border(Estonia,Estonia)

  (ii) Error:

No Error.

A.12. Author Relation

- Well-known German authors include Johann Wolfgang von Goethe, Friedrich Schiller, Gotthold Ephraim Lessing and Theodor Fontane.

  (i) Output:

author(Johann)

author(Wolfgang)

author(Goethe)

author(Friedrich)

author(Schiller)

author(Gotthold)

author(Ephraim)

author(Lessing)

author(Theodor)

author(Fontane)

(ii) Error:

No Error.

- Other major authors of that century include Alexandre Dumas (The Three Musketeers and The Count of Monte-Cristo), Jules Verne (Twenty Thousand Leagues Under the Sea), Émile Zola (Les Rougon-Macquart), Honoré de Balzac (La Comédie humaine), Guy de Maupassant, Théophile Gautier and Stendhal (The Red and the Black, The Charterhouse of Parma), whose works are among the most well known in France and the world.

(i) Output:

author(Alexandre)

author(Dumas)

author(Jules)

author(Verne)

author(Zola)

author(-LRB-/OLes)

author(Honoré)

author(Balzac)

author(Comédie)

author(Maupassant)

author(Gautier)

author(Stendhal)

(ii) Error*:*

No Error.

- In the 19th century the most popular composers were: Józef Elsner and his pupils Fryderyk Chopin and Ignacy Dobrzyński.

(i) Output:

author(Józef)

author(Elsner)

author(Chopin)

author(Ignacy)

author(Dobrzyński)

(ii) Error:

No Error.

- Among the most famous writers of the post-Soviet period are Oksana Zabuzhko, Yurii Andrukhovych, Oleksandr Irvanets (uk), Serhiy Zhadan, Taras Prokhasko, Jaroslav Melnik, Yuriy Izdryk (uk), Yuriy Pokalchuk, Yuriy Vynnychuk, Andrey Kurkov.

(i) Output:

author(Oksana)

author(Zabuzhko)

author(Yurii)

author(Andrukhovych)

author(Oleksandr)

author(Serhiy)

author(Zhadan)

author(Taras)

author(Prokhasko)

author(Jaroslav)

author(Melnik)

author(Yuriy)

author(Izdryk)

author(Yuriy)

author(Pokalchuk)

author(Yuriy)

author(Vynnychuk)

author(Andrey)

author(Kurkov)

(ii) Error:

No Error.

- Other writers from this period are Gustavo Adolfo Bécquer, José de Espronceda, Rosalía de Castro or Mariano José de Larra.

  (i) Output:

  author(Gustavo)
  author(Adolfo)
  author(Bécquer)
  author(José)
  author(Espronceda)
  author(Rosalía)
  author(Castro)
  author(Mariano)
  author(José)
  author(Larra)

  (ii) Error:

  No Error.

- Portuguese literature is represented by authors such as Almeida Garrett, Camilo Castelo Branco, Eça de Queiroz, Fernando Pessoa, Sophia de Mello Breyner Andresen, António Lobo Antunes and Miguel Torga.

  (i) Output:

  author(Almeida)
  author(Garrett)
  author(Camilo)
  author(Castelo)
  author(Branco)
  author(Eça)
  author(Queiroz)
  author(Fernando)
  author(Pessoa)
  author(Sophia)
  author(Mello)
  author(Breyner)

author(Andresen)
author(António)
author(Lobo)
author(Antunes)
author(Miguel)
author(Torga)

(ii) Error:

No Error.

- Japan has two Nobel Prize-winning authors Yasunari Kawabata (1968) and Kenzaburō Ōe (1994).

(i) Output:

author(Yasunari)
author(Kawabata)
author(Kenzaburō)

(ii) Error:

No Error.

- Historically, Iranian literature has inspired writers including Johann Wolfgang von Goethe, Henry David Thoreau, and Ralph Waldo Emerson.

(i) Output:

author(Johann)
author(Wolfgang)
author(Goethe)
author(Henry)
author(David)
author(Thoreau)
author(Ralph)
author(Waldo)
author(Emerson)

(ii) Error:

No Error.

- Egyptian women writers include Nawal El Saadawi, well known for her feminist activism, and Alifa Rifaat who also writes about women and tradition.

  (i) Output:

  author(Nawal)
  author(El)
  author(Saadawi)
  author(Alifa)
  author(Rifaat)

  (ii) Error:

  No Error.

- Major writers and novelists include Rómulo Gallegos, Teresa de la Parra, Arturo Uslar Pietri, Adriano González León, Miguel Otero Silva, and Mariano Picón Salas.

  (i) Output:

  author(Rómulo)
  author(Gallegos)
  author(Teresa)
  author(Parra)
  author(Arturo)
  author(Uslar)
  author(Pietri)
  author(Adriano)
  author(González)
  author(León)
  author(Miguel)
  author(Otero)
  author(Silva)
  author(Mariano)
  author(Picón)
  author(Salas)

(ii) Error:

No Error.

A.13. Composer Relation

- In the 19th century the most popular composers were: Józef Elsner and his pupils Fryderyk Chopin and Ignacy Dobrzyński.

  (i) Output:

  composer(Józef)
  composer(Elsner)
  composer(Chopin)
  composer(Ignacy)
  composer(Dobrzyński)

  (ii) Error:

  No Error.

- Italy's most famous composers include the Renaissance composers Palestrina and Monteverdi, the Baroque composers Scarlatti, Corelli and Vivaldi, the Classical composers Paganini and Rossini, and the Romantic composers Verdi and Puccini. Modern Italian composers such as Berio and Nono proved significant in the development of experimental and electronic music.

  (i) Output:

  composer(Monteverdi)
  composer(Scarlatti)
  composer(Corelli)
  composer(Vivaldi)
  composer(Paganini)
  composer(Rossini)
  composer(Verdi)
  composer(Puccini)
  composer(Berio)
  composer(Nono)

(ii) Error:

No Error.

- Dieterich Buxtehude composed oratorios for organ, which influenced the later work of Johann Sebastian Bach and Georg Friedrich Händel; these men were influential composers of the Baroque period.

(i) Output:

composer(Dieterich)
composer(Buxtehude)
composer(Johann)
composer(Sebastian)
composer(Bach)
composer(Georg)
composer(Friedrich)

(ii) Error:

No Error.

- The most renowned composers of this period include Marc-Antoine Charpentier, François Couperin, Michel-Richard Delalande, Jean-Baptiste Lully and Marin Marais, all of them composers at the court.

(i) Output:

composer(Marc-Antoine)
composer(Charpentier)
composer(François)
composer(Couperin)
composer(Michel-Richard)
composer(Delalande)
composer(Jean-Baptiste)
composer(Lully)
composer(Marin)
composer(Marais)

(ii) Error:

No Error.

- From traditional folk music, to classical and modern rock, Ukraine has produced several internationally recognised musicians including Kirill Karabits, Okean Elzy and Ruslana.

(i) Output:

composer(Kirill)
composer(Karabits)
composer(Okean)
composer(Elzy)
composer(Ruslana)

(ii) Error:

No Error.

- Music in 19th century Russia was defined by the tension between classical composer Mikhail Glinka along with other members of The Mighty Handful, who embraced Russian national identity and added religious and folk elements to their compositions, and the Russian Musical Society led by composers Anton and Nikolay Rubinsteins, which was musically conservative.

(i) Output:

composer(Mikhail)
composer(Glinka)
composer(Anton)
composer(Nikolay)
composer(Rubinsteins)

(ii) Error:

No Error.

- Augsburg-born composer Leopold Mozart mentored one of the most noted musicians of all time: Wolfgang Amadeus Mozart. Ludwig van Beethoven was a crucial figure in the transition between the Classical and Romantic eras.

(i) Output:

composer(Leopold)
composer(Mozart)
composer(Wolfgang)
composer(Mozart)
composer(Ludwig)
composer(Beethoven)

(ii) Error:

No Error.

- World-renowned composers of the 20th century include Alexander Scriabin, Igor Stravinsky, Sergei Prokofiev, Dmitri Shostakovich and Alfred Schnittke.

(i) Output:

composer(Alexander)
composer(Scriabin)
composer(Igor)
composer(Stravinsky)
composer(Sergei)
composer(Prokofiev)
composer(Dmitri)
composer(Shostakovich)
composer(Alfred)
composer(Schnittke)

(ii) Error:

No Error.

- The most renowned composers of this period include Marc-Antoine Charpentier, François Couperin, Michel-Richard Delalande, Jean-Baptiste Lully and Marin Marais, all of them composers at the court.

(i) Output:

composer(Marc-Antoine)
composer(Charpentier)

composer(François)
composer(Couperin)
composer(Michel-Richard)
composer(Delalande)
composer(Jean-Baptiste)
composer(Lully)
composer(Marin)
composer(Marais)

(ii) Error:

No Error.

- French composers from the Romantic era included: Hector Berlioz (best known for his Symphonie fantastique), Georges Bizet (best known for Carmen, which has become one of the most popular and frequently performed operas), Gabriel Fauré (best known for his Pavane, Requiem, and nocturnes), Charles Gounod (best known for his Ave Maria and his opera Faust), Jacques Offenbach (best known for his 100 operettas of the 1850s–1870s and his uncompleted opera The Tales of Hoffmann), Édouard Lalo (best known for his Symphonie espagnole for violin and orchestra and his Cello Concerto in D minor), Jules Massenet (best known for his operas, of which he wrote more than thirty, the most frequently staged are Manon (1884) and Werther (1892)) and Camille Saint-Saëns (he has many frequently-performed works, including The Carnival of the Animals, Danse macabre, Samson and Delilah (Opera), Introduction and Rondo Capriccioso, and his Symphony No. 3 (Organ Symphony)).

(i) Output:

composer(Hector)
composer(Berlioz)
composer(Georges)
composer(Bizet)
composer(Carmen)
composer(Gabriel)
composer(Fauré)
composer(Charles)

composer(Gounod)

composer(Maria)

composer(Faust-RRB-/O)

composer(Jacques)

composer(Offenbach)

composer(Hoffmann-RRB-/O)

composer(Lalo)

composer(Jules)

composer(Massenet)

composer(Manon)

composer(Camille)

composer(Saint-Saëns)

composer(Danse)

composer(Samson)

composer(Delilah)

composer(Rondo)

composer(Capriccioso)

(ii) Error:

No Error.

A.14. Artist Relation

- In Italy artists like Paolo Uccello, Fra Angelico, Masaccio, Piero della Francesca, Andrea Mantegna, Filippo Lippi, Giorgione, Tintoretto, Sandro Botticelli, Leonardo da Vinci, Michelangelo Buonarroti, Raphael, Giovanni Bellini, and Titian took painting to a higher level through the use of perspective, the study of human anatomy and proportion, and through their development of an unprecedented refinement in drawing and painting techniques.

  (i) Output:

artist(Paolo)

artist(Uccello)

artist(Fra)

artist(Masaccio)
artist(Piero)
artist(Francesca)
artist(Andrea)
artist(Mantegna)
artist(Filippo)
artist(Lippi)
artist(Giorgione)
artist(Tintoretto)
artist(Sandro)
artist(Botticelli)
artist(Leonardo)
artist(Vinci)
artist(Michelangelo)
artist(Buonarroti)
artist(Raphael)
artist(Giovanni)
artist(Bellini)
artist(Titian)

(ii) Error:

No Error.

- French artists developed the rococo style in the 18th century, as a more intimate imitation of old baroque style, the works of the court-endorsed artists Antoine Watteau, François Boucher and Jean-Honoré Fragonard being the most representative in the country.

  (i) Output:

artist(Antoine)
artist(Watteau)
artist(François)
artist(Boucher)
artist(Jean-Honoré)

artist(Fragonard)

(ii) Error:

No Error.

- Albrecht Dürer, Hans Holbein the Younger, Matthias Grünewald and Lucas Cranach the Elder were important German artists of the Renaissance, Peter Paul Rubens and Johann Baptist Zimmermann of the Baroque, Caspar David Friedrich and Carl Spitzweg of Romanticism, Max Liebermann of Impressionism and Max Ernst of Surrealism.

(i) Output:

artist(Albrecht)
artist(Dürer)
artist(Hans)
artist(Holbein)
artist(Matthias)
artist(Grünewald)
artist(Lucas)
artist(Peter)
artist(Paul)
artist(Rubens)
artist(Johann)
artist(Zimmermann)
artist(Caspar)
artist(David)
artist(Friedrich)
artist(Carl)
artist(Spitzweg)
artist(Max)
artist(Liebermann)
artist(Max)
artist(Ernst)

(ii) Error:

No Error.

- Distinguished contemporary artists include Roman Opałka, Leon Tarasewicz, Jerzy Nowosielski, Wojciech Siudmak, Mirosław Bałka, and Katarzyna Kozyra and Zbigniew Wąsiel in the younger generation.

(i) Output:

artist(Opałka)
artist(Leon)
artist(Tarasewicz)
artist(Jerzy)
artist(Nowosielski)
artist(Wojciech)
artist(Siudmak)
artist(Mirosław)
artist(Bałka)
artist(Katarzyna)
artist(Kozyra)
artist(Zbigniew)
artist(Wąsiel)

(ii) Error:

No Error.

- Famous ukiyo-e artists include Hokusai and Hiroshige.

(i) Output:

artist(Hokusai)
artist(Hiroshige)

(ii) Error:

No Error.

- The later "Group D" of artists led by Abidin Dino, Cemal Tollu, Fikret Mualla, Fahrünnisa Zeid, Bedri Rahmi Eyüboğlu, Adnan Çoker and Burhan Doğançay

introduced some trends that had lasted in the West for more than three decades.

(i) Output:

artist(Abidin)
artist(Dino)
artist(Cemal)
artist(Tollu)
artist(Fikret)
artist(Mualla)
artist(Fahrünnisa)
artist(Zeid)
artist(Bedri)
artist(Rahmi)
artist(Eyüboğlu)
artist(Adnan)
artist(Çoker)
artist(Burhan)
artist(Doğançay)

(ii) Error:

No Error.

- Notable Venezuelan artists include Arturo Michelena, Cristóbal Rojas, Armando Reverón, Manuel Cabré; the kinetic artists Jesús Soto, Gego and Carlos Cruz-Díez; and contemporary artists as Marisol and Yucef Merhi.

(i) Output:

artist(Arturo)
artist(Michelena)
artist(Cristóbal)
artist(Rojas)
artist(Armando)
artist(Reverón)
artist(Manuel)
artist(Jesús)

artist(Soto)

artist(Gego)

artist(Carlos)

artist(Cruz-Díez;/O)

artist(Marisol)

artist(Yucef)

artist(Merhi)

(ii) Error:

No Error.

- Post-revolutionary art in Mexico had its expression in the works of renowned artists such as David Alfaro Siqueiros, Federico Cantú Garza, Frida Kahlo, Juan O'Gorman, José Clemente Orozco, Diego Rivera, and Rufino Tamayo.

(i) Output:

artist(David)

artist(Alfaro)

artist(Siqueiros)

artist(Federico)

artist(Cantú)

artist(Garza)

artist(Frida)

artist(Kahlo)

artist(Juan)

artist(O'Gorman)

artist(José)

artist(Clemente)

artist(Orozco)

artist(Diego)

artist(Rivera)

artist(Rufino)

artist(Tamayo)

(ii) Error:

No Error.

- Notable artists from this era include El Lissitzky, Kazimir Malevich, Wassily Kandinsky, and Marc Chagall.

(i) Output:

artist(El)
artist(Lissitzky)
artist(Kazimir)
artist(Malevich)
artist(Wassily)
artist(Kandinsky)
artist(Marc)
artist(Chagall)

(ii) Error:

No Error.

- More recently, Colombian artists Pedro Nel Gómez and Santiago Martínez Delgado started the Colombian Murial Movement in the 1940s, featuring the neoclassical features of Art Deco.

(i) Output:

artist(Pedro)
artist(Nel)
artist(Gómez)
artist(Martínez)
artist(Delgado)
artist(Murial)
artist(Deco)

(ii) Error:

No Error.

A.15. Brand Relation

- Well-known international brands include Mercedes-Benz, BMW, SAP, Volkswagen, Audi, Siemens, Allianz, Adidas, Porsche, and DHL.

  (i) Output:

  brand(Mercedes-Benz)
  brand(Volkswagen)
  brand(Audi)
  brand(Siemens)
  brand(Adidas)
  brand(DHL)
  brand(Allianz)

  (ii) Error:

  No Error.

- The new car market is dominated by domestic brands such as Renault (27% of cars sold in France in 2003), Peugeot (20.1%) and Citroën (13.5%).

  (i) Output:

  brand(Renault)
  brand(Peugeot)
  brand(Citroën)

  (ii) Error:

  No Error.

- Well known Polish brands include, among others, PKO BP, PKN Orlen, PGE, PZU, PGNiG, Tauron Group, Lotos Group, KGHM Polska Miedź, Asseco, Plus, Play, PLL LOT, Poczta Polska, PKP, Biedronka, and TVP.

  (i) Output:

  brand(KGHM)
  brand(Orlen)
  brand(Tauron)
  brand(Lotos)

brand(Miedź)

brand(Asseco)

brand(Poczta)

brand(Biedronka)

(ii) Error:

No Error.

- Turkish brands like Beko and Vestel are among the largest producers of consumer electronics and home appliances in Europe, and invest a substantial amount of funds for research and development in new technologies related to these fields.

  (i) Output:

  brand(Vestel)

  brand(Beko)

  (ii) Error:

  No Error.

- Well-known Ukrainian brands include Naftogaz Ukrainy, AvtoZAZ, PrivatBank, Roshen, Yuzhmash, Nemiroff, Motor Sich, Khortytsa, Kyivstar and Aerosvit.

  (i) Output:

  brand(Naftogaz)

  brand(Roshen)

  brand(Yuzhmash)

  brand(Nemiroff)

  brand(Sich)

  brand(Khortytsa)

  brand(Aerosvit)

  (ii) Error:

  No Error.

A.16. Physicist Relation

- Galileo Galilei (1564–1642), a physicist, mathematician and astronomer, played a major role in the Scientific Revolution.

  (i) Output:

  physicist(Galileo)
  physicist(Galilei)

  (ii) Error:

  No Error.

- Physicist Enrico Fermi (1901–54), a Nobel prize laureate, led the team in Chicago that developed the first nuclear reactor and is also noted for his many other contributions to physics, including the co-development of the quantum theory and was one of the key figures in the creation of the nuclear weapon.

  (i) Output:

  physicist(Enrico)
  physicist(Fermi)

  (ii) Error:

  No Error.

- He, Emilio G. Segrè, and a number of Italian physicists were forced to leave Italy in the 1930s by Fascist laws against Jews, including Emilio G. Segrè (1905–89) (who discovered the elements technetium and astatine, and the antiproton), and Bruno Rossi (1905–93), a pioneer in Cosmic Rays and X-ray astronomy.

  (i) Output:

  physicist(Emilio)
  physicist(Segrè)
  physicist(Emilio)
  physicist(Segrè)
  physicist(Bruno)
  physicist(Rossi)

(ii) Error:

No Error.

- Other prominent physicists include: Amedeo Avogadro (most noted for his contributions to molecular theory, in particular the Avogadro's law and the Avogadro constant), Evangelista Torricelli (inventor of barometer), Alessandro Volta (inventor of electric battery), Guglielmo Marconi (inventor of radio), Ettore Majorana (who discovered the Majorana fermions), Carlo Rubbia (1984 Nobel Prize in Physics for work leading to the discovery of the W and Z particles at CERN).

(i) Output:

physicist(Amedeo)
physicist(Avogadro)
physicist(Avogadro's/O)
physicist(Avogadro)
physicist(Evangelista)
physicist(Torricelli)
physicist(Alessandro)
physicist(Volta)
physicist(Guglielmo)
physicist(Marconi)
physicist(Ettore)
physicist(Majorana)
physicist(Majorana)
physicist(Carlo)
physicist(Rubbia)

(ii) Error:

No Error.

- Notable German physicists before the 20th century include Hermann von Helmholtz, Joseph von Fraunhofer and Gabriel Daniel Fahrenheit, among others.

(i) Output:

physicist(Hermann)

physicist(Helmholtz)
physicist(Joseph)
physicist(Fraunhofer)
physicist(Gabriel)
physicist(Daniel)

(ii) Error:

No Error.

- Physicist Lene Vestergaard Hau is the first person to stop light, leading to advances in quantum computing, nanoscale engineering and linear optics.

(i) Output:

physicist(Lene)
physicist(Vestergaard)
physicist(Hau)

(ii) Error:

No Error.

- Internationally recognised Norwegian scientists include the mathematicians Niels Henrik Abel, Sophus Lie and Atle Selberg, physical chemist Lars Onsager, physicist Ivar Giaever, chemists Odd Hassel, Peter Waage, and Cato Maximilian Guldberg.

(i) Output:

physicist(Niels)
physicist(Henrik)
physicist(Abel)
physicist(Sophus)
physicist(Selberg)
physicist(Lars)
physicist(Onsager)
physicist(Ivar)
physicist(Giaever)
physicist(Hassel)
physicist(Peter)

physicist(Waage)

physicist(Cato)

physicist(Maximilian)

physicist(Guldberg)

(ii) Error:

Result is partly correct, as the names of the mathematicians, chemists and physicists are mixed.

Error Type: B

A.17. Mathematician Relation

- Galileo Galilei (1564–1642), a physicist, mathematician and astronomer, played a major role in the Scientific Revolution.

  (i) Output:

  mathematician(Galileo)

  mathematician(Galilei)

  (ii) Error:

  No Error.

- Fibonacci (c. 1170 – c. 1250), and Gerolamo Cardano (1501–76) made fundamental advances in mathematics.

  (i) Output:

  mathematician(Gerolamo)

  mathematician(Cardano)

  (ii) Error:

  Result is partly correct, as Fibonacci is missing.

  Error Type: B

- Numerous mathematicians were born in Germany, including Carl Friedrich Gauss, David Hilbert, Bernhard Riemann, Gottfried Leibniz, Karl Weierstrass, Hermann Weyl and Felix Klein.

(i) Output:

mathematician(Carl)
mathematician(Friedrich)
mathematician(Gauss)
mathematician(David)
mathematician(Hilbert)
mathematician(Bernhard)
mathematician(Riemann)
mathematician(Gottfried)
mathematician(Leibniz)
mathematician(Karl)
mathematician(Weierstrass)
mathematician(Hermann)
mathematician(Weyl)
mathematician(Felix)
mathematician(Klein)

(ii) Error:

No Error.

- The 18th-century priest, botanist and mathematician José Celestino Mutis was delegated by Viceroy Antonio Caballero y Góngora to conduct an inventory of the nature of the New Granada.

(i) Output:

mathematician(José)
mathematician(Celestino)
mathematician(Mutis)
mathematician(Viceroy)
mathematician(Antonio)
mathematician(Caballero)
mathematician(Góngora)
mathematician(Granada)

(ii) Error:

No Error.

- In addition to them, biologists Gregor Mendel and Konrad Lorenz as well as mathematician Kurt Gödel and engineers such as Ferdinand Porsche and Siegfried Marcus were Austrians.

(i) Output:

mathematician(Gregor)
mathematician(Mendel)
mathematician(Konrad)
mathematician(Lorenz)
mathematician(Kurt)
mathematician(Gödel)
mathematician(Ferdinand)
mathematician(Siegfried)
mathematician(Marcus)

(ii) Error:

Result is partly correct, as the names of biologists, engineers and mathematicians are mixed.

Error Type: B