DEBORA: Dependency-based Method for Extracting Entity-Relationship Triples from Open-Domain Texts in Polish

Alina Wróblewska[2], Marcin Sydow[1,2]

[1] Polish-Japanese Institute of Information Technology, Warsaw, Poland,
[2] Institute of Computer Science, Polish Academy of Sciences, Warsaw, Poland
alina@ipipan.waw.pl,msyd@poljap.edu.pl

Abstract. We present DEBORA – a dependency-based approach to the problem of extraction of arbitrary relations between named entities from open-domain texts in Polish. The presented method is designed for the purpose of the conducted experiment and is adapted to morpho-syntactic properties of Polish, e.g. free word order, high degree of morphological marking. Our preliminary results show that the method is applicable for Polish, even if there is a room for improvement. We also present the application of the extraction method in the problem of graphical entity summarisation.

Keywords: knowledge graphs, information extraction, natural language processing, dependency parser, graphical entity summarisation

1 Introduction

The amount of information available in textual resources on the Web is huge and growing. For this reason, *Information Extraction* (IE) that aims at automatic or semi-automatic collection of structured data of specific type from textual corpora of given domain (such as medicine, economics, etc.) is in the mainstream of academic and industrial research. More recently, IE scientists concentrate especially on *Open Domain* Information Extraction (ODIE), where information is automatically extracted from large textual resources, e.g. web news articles, which are not restricted to any particular domain or terminology.

In this paper we study a subtask of ODIE – the automatic *entity-relationship (ER) triple* extraction from Polish texts. ER-triples are instances of semantic binary relations between pairs of named entities, e.g. (*Warszawa, jest w, Polsce*), Eng. (*Warsaw, is located in, Poland*). Triples extracted from a text are treated as candidates for being facts. Thus, after applying some validation techniques to filter out invalid or unreliable facts, available ER-triples may be used to build a large semantic knowledge base concerning the real world. The focus of the paper is on the first stage of building a future knowledge base extracted from Polish open-domain textual corpora. This stage consists in extracting triples from Polish open-domain texts using a dependency-based technique.

Most of already proposed relation extraction techniques are based on pre-defined extraction rules or manually annotated training corpora. As the manual development of extraction patterns or the manual corpus annotation are expensive and time-consuming processes, systems based on these techniques are usually limited to one extraction domain. Recently, there is a growing interest in applying unsupervised or weakly supervised machine learning techniques to the relation extraction task, e.g. [6].

One of the first successful systems for the fast and scalable fact extraction from the Web is the domain-independent system, *KnowItAll* [4]. *KnowItAll* starts with the extraction of entities of pre-defined entity types (e.g. CITY, MOVIE) and then discovers instances of relations between extracted entities using handwritten patterns. Another system called *TextRunner* [2] applies a technique of extracting all meaningful instances of relations from the Web. The system *ReVerb* [5], in turn, overcomes some limitations of the mentioned systems due to a novel model of the verb-based relation extraction. *Enrycher* [10] is a system for triplets extraction and their visualisation in English and Slovene. The system applies a complex cascade of language processing tools, e.g., part-of-speech tagger, named entity recogniser, word sense disambiguation tool, parser identifying quasi predicate-argument structures, anaphora resolution and coreference resolution tools.

Although many efficient triple-extraction models exist for English and few other languages, this research field is still not explored in large group of inflectional languages with a relatively free word order, e.g., slavic languages. As was mentioned above, there exists a triplets extraction system for Slovene, however we could not find any information about the accuracy of extracted Slovene triplets. The direct application of many extraction techniques designed for English, which is an isolating language with topological argument marking, seems to be not suitable for inflectional languages with the morphological argument marking, such as Polish. Except for problems caused by rich morphology[1], topological extraction rules defined for English may not apply for a language with the free word order.

This paper presents preliminary experiments on extracting triples from Polish Web documents using a dependency-based method. To the best authors' knowledge, this is one of the first experiments dealing with the problem of open-domain triple extraction for Polish. We hope this publication will contribute to the discussion on this issue and to motivate further research.

The paper is structured as follows. Section 2 outlines the dependency-based method of extracting triples from Polish texts. The prototype implementation of the extraction heuristic is described in Section 3. Section 4 gives an overview of preliminary experimental results. Finally, we present a novel application of the extracted triples in graphical entity summarisation in Section 6.

[1] There is an adequacy between a case that marks a noun phrase and the argument this noun phrase may fulfill in Polish, e.g., a noun phrase marked for nominative is typically regarded as a subject in a sentence. But NPNOM may also be realised as a predicative complement, e.g. Pol. *Jan*$_{NOM}$ *to artysta*$_{NOM}$ (Eng. 'John is an artist.'). What is more, the Polish language is characterised by syncretism of forms, i.e., a single form may fulfill different grammatical functions, e.g., Pol. *Autobus*$_{NOM}$*-SUBJ wyprzedził samochód*$_{ACC}$*-OBJ* (Eng. 'The bus overtook a car.') vs. Pol. *Autobus*$_{ACC}$*-OBJ wyprzedził samochód*$_{NOM}$*-SUBJ* (Eng. 'The car overtook a bus.').

2 DEBORA – a Dependency-based Method of Triple Extraction

Although the existing triple extraction techniques may be efficient for English, they may not be applicable for other languages such as Polish. English is a language with a relatively restrictive word order used to convey grammatical information. Polish, in turn, is characterised by the high degree of morphological marking and rather flexible word order. Thus, a single fact may have numerous surface representations in a text.

Because of this, the iterative pattern induction as in *DIPRE* [3] or extraction of all meaningful facts defined as token chains between entities as in *TextRunner* [2] might be difficult or even inapplicable for Polish. As we do not have any manually specified domain-independent extraction patterns or seed instances of relations to start extraction of further facts, we aim at discovering triples using a dependency-based method.

A triple is defined as a tuple $t = (ne1_{subj}, r, ne2)$, where $ne1_{subj}$ denotes a nominal phrase recognised as a named entity fulfilling the subject function, $ne2$ represents another recognised named entity and r denotes an instance of a relation between these named entities. We decide to discover instances of relations only between recognised named entities, one of which fulfils the subject function. This decision is motivated by a property of the Polish language, which allows pro-drop pronouns with the subject function. At the current stage of our work, we do not want to model relations between implicitly realized entities. Sentences without a subject are ruled out, in order to avoid the coreference resolution problem, as we are not aware of any publicly available coreference resolutions tools for Polish.

Polish is a free word order language, so we may not rely on the order of named entities in a sentence while extracting triples. An example in Figure 1 shows that the identification of grammatical functions of named entities seems to be essential to extract meaningful instances of relations.[2]

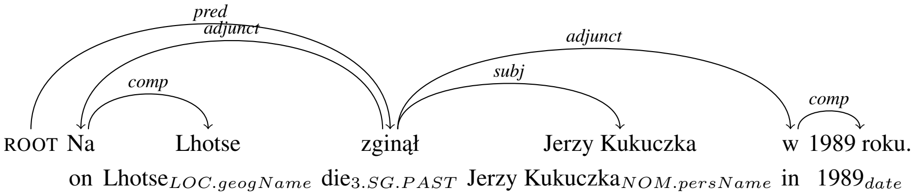

pred
adjunct
comp
adjunct
subj
comp
ROOT Na Lhotse zginął Jerzy Kukuczka w 1989 roku.
on Lhotse$_{LOC.geogName}$ die$_{3.SG.PAST}$ Jerzy Kukuczka$_{NOM.persName}$ in 1989$_{date}$

Fig. 1. The dependency structure of the Polish sentence *Na Lhotse zginął Jerzy Kukuczka w 1989 roku.* Eng. ‘Jerzy Kukuczka died 1989 on Lhotse.’

In our approach, only elements of the predicate-argument structure and adjuncts recognised as named entities are selected from the entire dependency structure. The relation between two named entities consists of a sentence predicate and its all arguments excluding those fulfilled by the two named entities. Currently, the field that triples are extracted from is restricted to a simple sentence or a matrix clause in a complex sentence, i.e. relations between named entities placed in different clauses are not regarded.

[2] Taking into account the topicalised part of a sentence, constituents in the example sentence in Figure 1 may be ordered differently, e.g. *Jerzy Kukuczka zginął na Lhotse w 1989 roku.*, *W 1989 roku na Lhotse zginął Jerzy Kukuczka.*, etc.

In an ideal scenario, a recognised named entity realised as a nominal phrase depends on the sentence predicate. This named entity takes part in the triple extraction. However, we also want to take into account named entities realised as noun phrases depending on a preposition (see preposition phrases '*Na Lhotse*' and '*w 1989 roku*' in Figure 1)[3] or another noun phrase (e.g. apposition) which are headed by the sentence predicate. These named entities are also involved in the extraction of instances of relations.

According to the above assumptions, two triples should be extracted from the dependency structure of the sentence shown in Figure 1:

- *zginął_na(Jerzy_Kukuczka*$_{NOM:persName:subj}$, *Lhotse*$_{LOC:geogName:comp}$*)*
- *zginął_w(Jerzy_Kukuczka*$_{NOM:persName:subj}$, *1989_roku*$_{LOC:date:comp}$*)*

3 Experimental Platform

The dependency-based triple extraction technique described above has been implemented and integrated with ExPLORER – a prototype experimental platform, being currently under development, for extracting fact database in the form of semantic knowledge base from open-domain texts in Polish.

ExPLORER can be generally viewed as a chain of configurable modules. It takes open-domain textual resources in Polish (e.g. web documents) as input and produces a set of extracted triples (optionally, enriched with some morpho-syntactic features) as output. Modules in the chain can be divided into the following functional groups:

- (optional) focused web crawling module
- text extraction (filtering out non-text elements such as pictures, ads, navigation; removing html tags, etc.)
- NLP processing (POS tagging, lemmatisation, NER, dependency parsing)
- ER triples extraction

In short, the processing is as follows: a large open-domain textual corpus is automatically crawled, analysed and annotated with external language processing tools. Then, instances of relations between two named entities are automatically extracted.

Regarding NLP-processing, we use publicly available, but largely imperfect tools for Polish. We start with the best Polish part-of-speech tagger *Pantera* [1], which divides the entire text into sentences and tokens, performs a thorough morphological analysis and augments tokens with their lemmas, part-of-speech tags and morpho-syntactic features. Next, morpho-syntactically annotated texts are given as an input to a named-entity recogniser *Nerf*[4] [8], which annotates personal names, organisation names, place names and dates. The corpus annotated with morpho-syntactic features and named entities constitutes an input to the Polish dependency parser *MaltParser*s[5] [11]. The NLP-processing phase may be particularly prone to errors, due to using imperfect NLP tools.

Finally, instances of relations between two named entities are automatically extracted with an extraction module based on the heuristic described in Section 2.

[3] If a noun phrase recognised as a named entity is governed by a preposition, the preposition in appended to the instance of relation.

[4] According to [8], *Nerf* achieves the general recognition performance of 79% F-score.

[5] According to [11], the Polish MaltParser achieves the parsing performance of 71% *labelled attachment score*.

4 Experimental Results

The extraction method described in section 2 is applied to a set of Polish web news articles (188,415 texts) taken from [7]. Raw texts are split into 6,303,794 sentences with 20.3 tokens per sentence on average. As the goal of the experiment is to discover triples relating named entities, only these sentences with at least one recognised named entity (3,265,817 sentences) are parsed with the Polish dependency parser. The morphosyntactically annotated and dependency-parsed sentences are given to the triple extractor that discovers 58,742 instances of relations between pairs of named entities. The extracted triples concerned 26,469 unique named entities fulfilling the subject function.In order to evaluate the quality of extracted triples and the extraction procedure itself, two evaluations are carried out.

4.1 First Evaluation Experiment

Because the total number of extracted triples is quite large in our web-based experiment, the straightforward computation of *precision* is not a trivial task. Furthermore, the exact computation of *recall* in case of a large web-based input text corpus is completely infeasible, since it would involve counting all valid triples contained in this corpus. Instead, we compute an approximation of precision by sampling 100 random triples and manually examining their validity. We repeat this computation three times and achieve the average precision of 54% (see Table 1, columns 2–4).
We also select all triples which represent some particularly interesting (from the application point of view) relations concerning people and places such as *bornIn, died, livedIn, isLocatedIn*. The precision of these relations is also manually computed (see Table 1, columns 5–8).

Table 1. Precision for selected samples of relations extracted from the Polish web news articles

relation:	random 100 a	random 100 b	random 100 c	bornIn	died	livedIn	isLocatedIn
total:	100	100	100	154	228	16	7
precision:	55%	53%	54%	98.7%	80%	87.5%	100%

The results are very promising. Despite the early stage of our work and difficulty with the open-domain extraction task (especially for Polish), the majority of the examined extracted triples are correctly formed and represent interesting facts about entities (see Figure 2).

4.2 Second Evaluation Experiment

The triple extraction process may be prone to errors, as imperfect NLP-tools are used. The second evaluation is performed, in order to check the impact of the linguistic processing on the quality of extracted triples.

For reasons of this evaluation, we prepared a small text in Polish to test the DEBORA algorithm. The text is a concatenation of three short biographical notes on Maria

Adam_Mickiewicz urodził_się_w Zaosiu (Eng. "born in" (location))
Mickiewicz urodził_się 24_grudnia_1798_r_. (Eng. "born on" (date))
Adam_Mickiewicz urodził_się_W pobliżu_Nowogródka (Eng. "was born close to")
Mickiewicz przeżył_na_zesłaniu_w Rosji (Eng. "survived the exile to Russia")
Mickiewicz wierzył_W Boga (Eng. "believed in God")
Adam_Mickiewicz bywał_w Szczorsach (Eng. "used to be in" (location))

Fig. 2. Triples concerning Adam Mickiewicz – one of the greatest Polish poets in the 19th century – automatically extracted from the Polish web corpus.

Skłodowska-Curie (an outstanding Polish 20-th century double nobel-prize winning scientist: physics and chemistry), Jacek Malczewski (a famous Polish 19-th century painter), and Robert Lewandowski (a contemporary Polish soccer player, who scored a goal in the first match played by the Polish soccer team on EURO 2012). The text consists of 64 quite simple sentences in total, with 9.25 tokens per sentence on average.

Prior to the experiment we prepared the gold-truth set of 100 triples that are contained in the text in order to compute the evaluation measures for the extraction method.

An excerpt of the evaluation text is presented in the appendix.

In the first step of the evaluation experiment (baseline), 62 triples are automatically extracted from the text and 46 of them are correct (precision: 74.2%, recall: 46%) (see Table 2, second row). We find out that Nerf has not recognised any of six alone occured last names. Sentences with not recognised named entities are not taken into account while extracting triples.

That is why in the second step of the performed experiment, input given to the triple extractor is manually corrected, i.e. part-of-speech tags and dependency relations[6] are amended and some missing named entity labels are added. Manual corrections of input increase the number of extracted triples (96 correct triples, 93.9% of F-score, see Table 2, third row).

Table 2. Evaluation of automatically extracted triples against a set of 12 gold standard instances of relations. Evaluation metrics: precision, recall and F-measure. Explanation: POS & NER (manual) – manually corrected part-of-speech tagging and named entity recognition, DP (manual) – manually corrected dependency structures.

experiment	POS & NER (manual)	DP (manual)	correct triples	precision	recall	F-score
1 (baseline)	–	–	46	74.2%	46%	56.8%
2	+	+	92	95.8%	92%	93.9%

4.3 Discussion

The presented results of both experiments are quite successful. Even if DEBORA leaves much room for improvement on hard, open-domain texts crawled from the Polish web (54% of precision), it quite successfully extracts some particular important relations.

[6] A named entity may be incorrectly annotated as a dependent of a constituent instead of a sentence predicate or additional arguments are not subcategorised by a predicate

Furthermore, on clearer and simpler texts it performs much better (74.2% of precision) especially after providing hand-corrected, high-quality input to it (95.8% of precision).

The results also clearly show that while the described extraction method is applicable for Polish, better NLP-tools, especially the improved Polish dependency parser and NER-tool, are needed, in order to further improve the extraction performance.

We also perform a cursory error analysis indicating that many of the extracted triples may not be regarded as proper pattern candidates due to many reasons. The poor quality of some extracted triples is mainly the result of the error-prone linguistic processing and leaves much room for future improvement. Among the most typical errors, we observe: missing triples caused by unrecognised named entities, partially identified relation and incorrectly composed relation caused by errors in dependency structures.

According to [2], deploying a deep linguistic parser to extract relations between entities is not practicable at Web scale. We showed that it can be a reasonable solution to extract facts from open-domain unstructured texts in a morphologically rich language with the free word order, such as Polish. Except for these properties of the Polish language, we observed some other problems while extracting triples. First, Polish allows for implicit realisations of pro-drop pronouns with the subject function. Second, almost all constituent types may be omitted in Polish. As no coreference resolution tools or any ellipsis detector exist for Polish and we could not manage mentioned problems, the extraction of all possible instances of relations seems to be highly problematic. That is why we concentrated our attention on extraction instances of relations between explicitly recognised named entities. Instances of relations between two named entities consist of a sentence predicate and its all arguments excluding those fulfilled by the named entities.

5 Potential Future Applications of DEBORA to Graphical Entity Summarisation

Extracted triples, after additional post-processing (e.g. NE normalisation and disambiguation) may be used to automatically build large semantic knowledge bases that can be viewed as large repositories of facts automatically extracted from the open-domain sources like www. Such repositories can be further processed or queried.

As a demonstration of such future possibilities, we present an example of an application of DEBORA to compute *graphical entity summarisations* on semantic knowledge graphs [9].

Figure 3 presents a graphical summary of the Polish 19th century poet *Adam Mickiewicz* automaticall created with a diversified summarisation tool developed in the DIVERSUM project [9] applied to the set of triples concerning the poet automatically extracted by DEBORA from the Polish web corpus described in Section 4.

The presented example of graphical summary is of surprisingly high quality, especially when one takes into account that it is based on open-domain web articles.

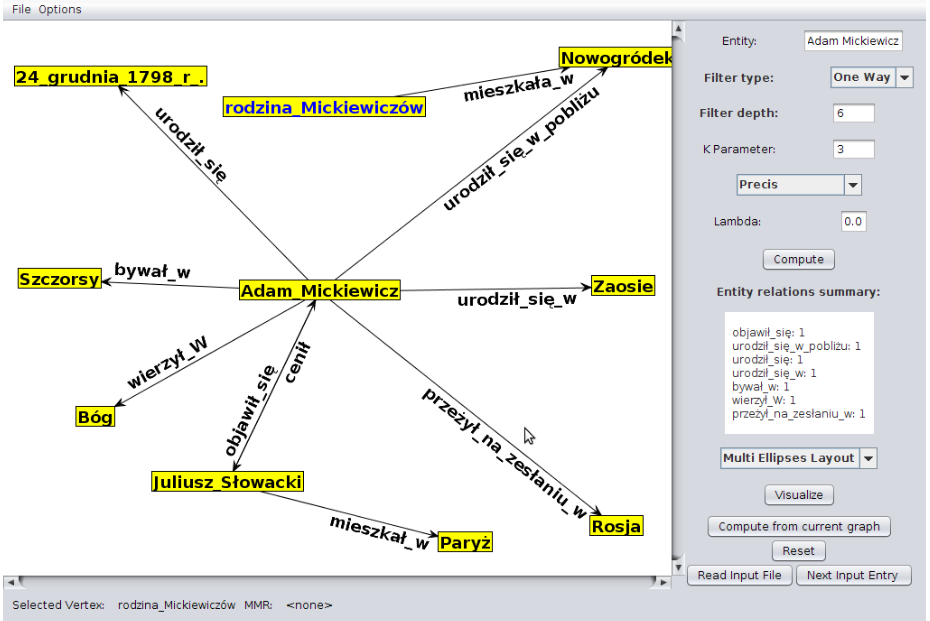
It would be very interesting to further integrate the crawling, extracting, summarising and visualising modules into one coherent platform.

One may imagine two operational modes of such platform.

In the off-line mode, the user first specifies the web sources to be automatically collected off-line by an intelligent focused web crawler. The crawled corpus is subsequently processed by DEBORA in the off-line manner to build a large knowledge graph that contains extracted facts on named entities from a given domain. Finally, such a knowledge base can be interactively queried by users with the tool similar to the one presented on Figure 3.

In the on-line mode, user provides the system with a medium-sized passage of text concerning some domain or entity (similarly to the biographical text used in the second evaluation experiment in Section 4.2). The text is immediately processed by DEBORA and user can interactively use the system to produce graphical summarisations of the entities concerned with the input text.

Since similar systems already exist for English and some other languages the authors are not aware of the existence of such for Polish. Our experimental platform under development, presented in this section, seems to be a promising prototype to achieve the described functionality for Polish.



Entity: Adam Mickiewicz
Filter type: One Way
Filter depth: 6
K Parameter: 3
Precis
Lambda: 0.0

```
objawił_się: 1
urodził_się_w_pobliżu: 1
urodził_się: 1
urodził_się_w: 1
bywał_w: 1
wierzył_W: 1
przeżył_na_zesłaniu_w: 1
```

Multi Ellipses Layout

Selected Vertex: rodzina_Mickiewiczów MMR: <none>

Fig. 3. Graphical entity summarisation obtained with a visualisation tool described in [9] concerning the Polish poet Adam Mickiewicz based on automatically extracted facts from Polish web texts. (NEs were normalised manually for this example)

6 Conclusions

We presented DEBORA – a method for extracting ER-triples from Polish open-domain texts based on a dependency parser for Polish. The method was implemented in an experimental ExPlorer platform and preliminarily evaluated on real data consisting of web documents as well as on prepared passages of texts.

Achieved results show that the dependency-based extraction method may be quite successfully applied to extract triples of Polish open-domain texts. According to results of the evaluation based on randomly selected samples of triples, we achieved an approximate precision of about 54%. Furthermore, an average precision of about 90% characterised selected triples representing some favourable relations such as *bornIn, died, livedIn isLocatedIn*. Experiments carried out in the second evaluation, which was based on a small test corpus, confirmed that the quality of the linguistic processing has a huge impact on the number and quality of extracted triples. There is no doubt that the better NLP-tools are at hand the better results might be achieved. Even if our current results may not be comparable with results of relation extraction achieved for English, they certainly encourage us to the further work.

In the future research, we are going to improve our extraction procedure and apply some filtering, normalisation and disambiguation techniques. We plan to annotate a large, heterogeneous Web corpus with the automatically extracted relations, in order to extend the triple set. Normalised and validated facts will be used to build a large semantic knowledge base for Polish.

Acknowledgements The work is a part of the DIVERSUM project hold at the Web Mining Lab of Polish-Japanese Institute of Information Technology and is supported by the N N516 481940 grant of National Science Centre.

References

1. Szymon Acedański. A morphosyntactic brill tagger for inflectional languages. In Hrafn Loftsson, Eiríkur Rögnvaldsson, and Sigrún Helgadóttir, editors, *Advances in Natural Language Processing*, volume 6233 of *Lecture Notes in Computer Science*, pages 3–14. Springer, 2010.
2. Michele Banko, Michael Cafarella, Stephen Soderland, Matt Broadhead, and Oren Etzioni. Open information extraction from the web. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, pages 2670–2676, 2007.
3. Sergey Brin. Extracting patterns and relations from the world wide web. In *Selected papers from the International Workshop on The World Wide Web and Databases*, pages 172–183, London, UK, 1999. Springer-Verlag.
4. Michael Cafarella, Doug Downey, Stephen Soderland, and Oren Etzioni. Knowitnow: Fast, scalable information extraction from the web. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2005.
5. Oren Etzioni, Anthony Fader, Janara Christensen, Stephen Soderland, and Mausam. Open information extraction: The second generation. In *Proceedings of the 20th International Joint Conference on Artifical Intelligence*, 2011.
6. Takaaki Hasegawa, Satoshi Sekine, and Ralph Grishman. Discovering relations among named entities from large corpora. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, ACL '04, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics.
7. Korpus rzeczpospolitej. [on-line] `http://www.cs.put.poznan.pl/dweiss/rzeczpospolita`.

8. Agata Savary and Jakub Waszczuk. Narzedzia do anotacji jednostek nazewniczych. In Adam Przepiórkowski, Mirosław Bańko, Rafał L. Górski, and Barbara Lewandowska-Tomaszczyk, editors, *Narodowy Korpus Języka Polskiego*. Wydawnictwo Naukowe PWN, Warsaw, 2012. Forthcoming.
9. Marcin Sydow, Mariusz Pikuła, and Ralf Schenkel. To diversify or not to diversify entity summaries on rdf knowledge graphs? In Marzena Kryszkiewicz, Henryk Rybinski, Andrzej Skowron, and Zbigniew Ras, editors, *Foundations of Intelligent Systems, Proc. of the 19th ISMIS Conference, Warsaw, Poland, 2011*, volume 6804 of *Lecture Notes in Artificial Intelligence*, pages 490–500. Springer Berlin / Heidelberg, 2011. 10.1007/978-3-642-21916-0-53.
10. Tadej Štajner, Delia Rusu, Lorand Dali, Blaž Fortuna, Dunja Mladeni/'c, and Marko Grobelnik. A service oriented framework for natural language text enrichment. *Informatica (Slovenia)*, 34(3):307–313, 2010.
11. Alina Wróblewska. Polish dependency bank. *Linguistic Issues in Language Technology*, 7(1), 2012.

Appendix

Below, there is an excerpt of the evaluation text concerning Maria Skłodowska-Curie used in the experiment reported in section 4.2. The parts concerning two other Polish figures are not shown here. For the ease of presentation here, it is separated into sentences with English translations coming after each sentence.

Maria Salomea Skłodowska-Curie urodziła się 7 listopada 1867 w Warszawie (Maria Salomea Skłodowska-Curie was born on 7.11.1867 in Warsaw)
Skłodowska-Curie zmarła 4 lipca 1934 w Passy (Skłodowska-Curie died on 4.07.1934 in Passy)
Skłodowska-Curie była polsko-francuską uczoną (Skłodowska-Curie was a Polish-French scientist)
Skłodowska-Curie większość życia spędziła we Francji (Skłodowska-Curie spent most of her life in France)
Maria studiowała we Francji (Maria studied in France)
Skłodowska-Curie została dwukrotnie wyróżniona Nagrodą Nobla za osiągnięcia naukowe (Skłodowska-Curie was twice awarded by the Nobel Prize for her scientific achievements)
Pierwszą Nagrodę Nobla Skłodowska-Curie dostała w roku 1903 z fizyki wraz z mężem i Henrim Becquerelem za badania nad odkrytym przez Becquerela zjawiskiem promieniotwórczości (Her first Nobel Prize Skłodowska-Curie obtained in 1903 in physics together with her husband Henri Becquerel for their research on the radiation phenomenon discovered by Becquerel)
Po raz drugi Skłodowska-Curie została wyróżniona Nargrodą Nobla w 1911 roku z chemii za wydzielenie czystego radu i badanie właściwości chemicznych pierwiastków promieniotwórczych (Second time Skłodowska-Curie was awarded the Nobel Prize in 1911 in chemistry for separating pure Radium and studying chemical properties of radioactive elements)
Skłodowska-Curie była żoną Pierre'a Curie (Skłodowska-Curie was Pierre's Curie wife)
Maria Skłodowska rozpoczęła naukę na Sorbonie w listopadzie 1891 roku (Maria Skłodowska begun her studies on Sorbone in November 1891)
Nauczycielem Skłodowskiej był Paul Appel (Skłodowska's teacher was Paul AppelAppel)
Nauczycielem Skłodowskiej był Henri Poincaré (Skłodowska's teacher was Henri Poincaré)
Nauczycielem Skłodowskiej był Gabriel Lippmann (Skłodowska's teacher was Gabriel Lippmann)