A system for named entity recognition and semantic relation extraction: An application to Modern Greek biomedical texts

1. Introduction

The number of documents in the digital universe is rapidly increasing. The biomedical literature is no exception to this trend, making it difficult for researchers to stay abreast of biomedical knowledge. This domain contains a wide range of documents, among them full-text articles or abstracts published in journals, books, medical case reports, medical regulatory documents, manuals, pharmaceutical brochures or catalogues. It is therefore useful to research means of automatically extracting knowledge from these documents using Natural Language Processing techniques.

According to Sekine (2004), the notion Named Entity Recognition (NE), was born in the Message Understanding Conference (MUC)-6 in 1995 (Grishman & Sundheim 1996). In an attempt to respond to the needs imposed, it has been extended beyond proper names, numeric expressions and dates and includes terms as well as proper names or other words or expressions that comprise the biomedical vocabulary. The need for an extended Named Entity Recognition was already noted by Sekine (2004): “The named entity [recognition] task is changing from tagging only proper names to tagging a broader range of words and expressions that are of interest to people with particular information needs. In any case, the named entity or extended named entity task is definitely one of the important component technologies for the applications of natural language technologies.” As Zweigenbaum & Demner-Fushman (2009, 356) point out: “Interesting entities in the biomedical domain include genes, proteins, diseases, drugs, body parts, etc.”
2. Our research

The goal of our research is to recognize Greek biomedical named entities in texts and the semantic relations between them. Despite the fact that several research projects have already been conducted on this topic, existing studies concern mainly English or other widely spoken languages (Chernyshevich & Stankevitch 2015; Di Buono 2015; Kumar et al. 2015; Névéol et al. 2014; Wang et al. 2014). Biomedical documents written in Greek, which is a less spoken language, have not yet received the attention they merit and they have been either cursorily examined or totally ignored.

The method we implemented to recognize named entities is dictionary based, which means that the resources used must be as complete and detailed as possible. The dictionary was applied to the corpus described below.

3. Corpus collection

Our corpus contains the summaries of 986 European public assessment reports (EPAR) in Greek (available at http://www.ema.europa.eu/ema/). An EPAR is “A set of documents describing the evaluation of a medicine authorized via the centralized procedure and including the product information, published on the European Medicines Agency website. European public assessment reports include the product information.” Their language is in a user friendly question and answer format intended for human use and has great structural diversity. Documents were downloaded in PDF format, saved, and automatically converted to plain text files suitable for being processed and merged into one final corpus which amounted to 1,052,437 words.

4. UMLS Knowledge base and MetaMap

In order to ensure consistency with what is considered to be a named entity, we used the Unified Medical Language System (UMLS), “a set of files and software that bring together many health and biomedical vocabularies and standards to enable interoperability between computer systems” (available at https://uts.nlm.nih.gov/home.html). UMLS was created by the United States National Library of Medicine (NLM) and English is the predominant language used. The UMLS consists of three tools. The first is Metathesaurus, which is “a large biomedical thesaurus that is organized by concept, or meaning, and it links similar names for the same concept from nearly 200 different vocabularies. The Metathesaurus also identifies useful relationships between concepts and preserves the meanings, concept names, and relationships from each vocabulary.” The second is Semantic Network, which is “a set of broad subject categories, or Semantic Types, that provide a consistent categorization of all concepts represented in the UMLS Metathesaurus, and a set of useful and impor-

NAMED ENTITY RECOGNITION AND SEMANTIC RELATION EXTRACTION

tant relationships, or Semantic Relations, that exist between Semantic Types.”

The third is the Specialist Lexicon and Lexical Tools which provide tools that can facilitate Natural Language Processing. The semantic types provided by the UMLS can be clustered into 15 semantic groups (McCray, Burgun & Bodenreider 2001; Névéol et al. 2014). We chose to deal with the semantic types of ANATOMY and DISORDERS, two of the largest classes of concepts in the UMLS (McCray, Burgun & Bodenreider 2001). The semantic types grouped under the umbrella of ANATOMY are the following: Anatomical Structure, Body Location or Region, Body Part, Organ, or Organ Component, Body Space or Junction, Body Substance, Body System, Cell, Cell Component, Embryonic Structure, Fully Formed Anatomical Structure and Tissue. Similarly, DISORDERS contains the types Acquired Abnormality, Anatomical Abnormality, Cell or Molecular Dysfunction, Congenital Abnormality, Disease or Syndrome, Experimental Model of Disease, Finding, Injury or Poisoning, Mental or Behavioral Dysfunction, Neoplastic Process, Pathologic Function and Sign or Symptom.

A small corpus was used for experimentation and prototyping. Rather than process the entire dataset, which would be computationally expensive, a sample of 300 documents was randomly selected.

In order to identify named entities included in our corpus and match them correctly to the ones from the UMLS database, we subsequently downloaded the same sample of EPAR written in English and we followed the same procedure as we followed with the Greek documents to unify them into a single corpus.

We therefore built a parallel corpus:

<table>
<thead>
<tr>
<th></th>
<th>Words</th>
<th>KB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greek</td>
<td>331 367</td>
<td>3 731&lt;sup&gt;3&lt;/sup&gt;</td>
</tr>
<tr>
<td>English</td>
<td>296 403</td>
<td>4 374</td>
</tr>
</tbody>
</table>

To map our English corpus to concepts in the UMLS we used MetaMap (available at https://www.ncbi.nlm.nih.gov/pubmed/20442139). Our search was based on the 2015 AB release. We restricted our search by selecting the semantic types we were interested in and by the Concept Unique Identifier (CUI)<sup>4</sup> for each named entity. For

---

2 https://semanticnetwork.nlm.nih.gov/

3 The size of the data corresponds to Unicode encoding.

4 “The Concept Unique Identifier for a Metathesaurus entry describes a concept to which strings with the same meaning are linked. One of the principles of the Metathesaurus is that meanings should be preserved over time regardless of what terms (atoms) are used to express those meanings. The CUI is an identifier that uniquely represents a meaning and (ideally) over time the meaning of a CUI does not change. As sources are updated and as Metathesaurus editors discover errors or find that the meanings of terms have shifted over time, the meaning corresponding to a CUI may be altered (merged or split) or may disappear. In such cases, the changes in the meaning of the CUIs involved are tracked in the CUI history table, so that any CUI from any previous release of the Metathesaurus may always be mapped to the equivalent concept in the current Metathesaurus (if any) or may be identified as having been deleted (in which case the closest similar concept will often be specified)”. (UMLS glossary, available at https://www.nlm.nih.gov/research/umls/new_users/glossary.html)
instance, *Lung diseases* and *Pulmonary diseases* are listed under the same CUI which is C0024115. Likewise, *Headache* and *Cephalalgia* are both encoded under C0018681. We opted for the default output for mappings in order to display only the top-scoring ones (see Aronson 2001).

A variety of output formats were offered. An example of the human-readable output for the input *bleeding in the joints* is presented below. As it can be seen, the overall score is 783. The concept’s score, the UMLS string matched, the concept’s Preferred Name and the concept’s Semantic Type(s) are also provided.5

Phrase: such as bleeding in the joints
Meta Mapping (783):
783 C0018924:bleeding joints (Hemarthrosis) [Pathologic Function]

As a preliminary step, we first manually assessed the results we received using MetaMap. Clinical signs and symptoms can have the semantic type Finding. However, the converse statement is not always true: concepts mapped as a Finding are not always clinical signs or symptoms. In order to preserve the robustness of the results, we excluded from the scope of this research common verbs such as *used* or *read*, percentages or telephone numbers mapped as Findings.

Nevertheless, to optimize the results we had to take a number of steps described below. Named entities recognized can be one-word (e.g. *heart, blood, diabetes*) or multi-word (e.g. *heart attack, blood vessel, diabetes complication*). In the case of the latter, we decided to consider their constituent words as named entities as long as a corresponding concept was included in the UMLS. For instance, *Myocardial Infarction* (CUI: C0027051) is composed of *Myocardial* (CUI: C1522564) and *Infarction* (CUI: C0021308). We dealt in exactly the same way with sequences like *lung or heart problems*, connected with a coordinating conjunction. Some words were excluded because there were not covered in the UMLS.

5. Greek named entities

Documents in English and Greek were then processed by XAlign, an alignment tool integrated in Unitex.6 Alignment allowed us to match English named entities to their correspondent ones in Greek. After the completion of the automated procedure, we carried out a manual verification of the output to detect misalignments.

Lexical or grammatical words used in English named entities may be varied in relation to the ones used to form their respective Greek named entities, e.g. *episodes of depression* corresponds to *καταθλιπτικά επεισόδια* (Adjective + Noun) or *abdominal surgery* corresponds to *χειρουργική επέμβαση κοιλίας* (Adjective + Noun + Noun). Moreover, they can be either one-word or multi-word, e.g. *sinuses* corresponds to *παραρρίνιοι κόλποι* (Adjective + Noun), *heart disease* corresponds to *καρδιοπάθεια*

---

6 See Section 7.
A double structure is also a possibility: στομαχικός καύσος (Adjective + Noun) or καύσος στομάχου (Noun + Noun) (heartburn). Finally, there are English named entities that correspond to two synonymous Greek words, e.g. body is σώμα or οργανισμός.

The occurrences retrieved as regards each Semantic Type of the two Semantic Groups are summarized in the following tables:

### ANATOMY

<table>
<thead>
<tr>
<th>Semantic Type</th>
<th>Occurrences</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anatomical Structure</td>
<td>387</td>
<td>σώμα (body)</td>
</tr>
<tr>
<td>Body Location or Region</td>
<td>411</td>
<td>ώμος (shoulder)</td>
</tr>
<tr>
<td>Body Part, Organ, or Organ Component</td>
<td>2,046</td>
<td>εγκέφαλος (brain)</td>
</tr>
<tr>
<td>Body Space or Junction</td>
<td>129</td>
<td>στόμα (mouth)</td>
</tr>
<tr>
<td>Body Substance</td>
<td>930</td>
<td>αίμα (blood)</td>
</tr>
<tr>
<td>Body System</td>
<td>729</td>
<td>νευρικό σύστημα (nervous system)</td>
</tr>
<tr>
<td>Cell</td>
<td>954</td>
<td>αιμοπετάλιο (blood platelet)</td>
</tr>
<tr>
<td>Cell Component</td>
<td>402</td>
<td>ριβοσώματα (ribosomes)</td>
</tr>
<tr>
<td>Embryonic Structure</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>Fully Formed Anatomical Structure</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>Tissue</td>
<td>357</td>
<td>μύς (muscle)</td>
</tr>
<tr>
<td><strong>Total</strong>: 6,345 occurrences</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### DISORDERS

<table>
<thead>
<tr>
<th>Semantic Type</th>
<th>Occurrences</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquired Abnormality</td>
<td>105</td>
<td>τόφος (tophus)</td>
</tr>
<tr>
<td>Anatomical Abnormality</td>
<td>33</td>
<td>παραρρίνιοι κόλποι (sinuses)</td>
</tr>
<tr>
<td>Cell or Molecular Dysfunction</td>
<td>30</td>
<td>ευθραυστότητα (fragility)</td>
</tr>
<tr>
<td>Congenital Abnormality</td>
<td>18</td>
<td>διαταραχές του οισοφάγου (abnormality of the esophagus)</td>
</tr>
<tr>
<td>Disease or Syndrome</td>
<td>4,056</td>
<td>υπερασβεστιαιμία (hypercalcaemia)</td>
</tr>
<tr>
<td>Experimental Model of Disease</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>Finding</td>
<td>4,782</td>
<td>συμπτώματα των ασθενών (patient symptoms)</td>
</tr>
<tr>
<td>Injury or Poisoning</td>
<td>348</td>
<td>κάταγμα (fracture)</td>
</tr>
<tr>
<td>Mental or Behavioral Dysfunction</td>
<td>651</td>
<td>σχιζοφρένεια (schizophrenia)</td>
</tr>
<tr>
<td>Neoplastic Process</td>
<td>1,059</td>
<td>καρκίνος (cancer)</td>
</tr>
<tr>
<td>Pathologic Function</td>
<td>1,977</td>
<td>σκληρία (induration)</td>
</tr>
<tr>
<td>Sign or Symptom</td>
<td>2,455</td>
<td>πόνος (pain)</td>
</tr>
<tr>
<td><strong>Total</strong>: 15,514 occurrences</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6. Problems

Despite the good performance of the UMLS and MetaMap, some shortcomings have already been observed in earlier similar researches (Okumura & Tateisi 2012, 69; Tran et al. 2015, 20). Problems hindering the task of named entity recognition include the following:

Non-recognition or misrecognition of named entities

In some cases, an adjective or a verb was not recognized, despite the fact that the corresponding noun was a concept, e.g. *hypersensitive* and [C0020517] Hypersensitivity [Pathologic Function] or *the bone that is naturally broken down* and [C1265875] Breakdown [Acquired Abnormality].

Another case where a named entity was absent from the results was when, despite the fact that it should be recognized as a Sign or Symptom, it was not. For instance, in the sentence “Parkinson’s disease is a progressive brain disorder that causes shaking, slow movement and muscle stiffness.”, only *shaking* and *muscle stiffness* were recognized.

In addition, other problems observed were due to erroneous lexical parsing, as some multi-word expressions were not recognized as an entire unit. For example, in *severe Alzheimer’s disease*, *Alzheimer* was not recognized and the result provided was C1836348:Severe Disease (Severe disorder) [Finding].

A similar problem occurred in the case of the phrase *soft contact lenses*:

Meta Mapping (802):

827  C3542022:SOFT (SHORT STATURE, ONYCHODYPLASIA, FACIAL DYSMORPHISM, AND HYPOTRICHOSIS SYNDROME) [Disease or Syndrome]
827  C0023317:LENS (Lens, Crystalline) [Body Part, Organ, or Organ Component]

Finally, false positives were obtained due to wrong phrase chunking. For example, in the sentence “Patients with mild liver problems should use Azilect with caution”, the sequence “patients with mild liver problems” was recognized as C1254481: patients problems (Patient Problem) [Finding] and C0023884:LIVER (Liver) [Body Part, Organ, or Organ Component], despite the fact that in the sequence “moderate liver problems” of the sentence “It is not recommended for patients with moderate liver problems.”, the phrase *liver problems* was correctly mapped as a Finding.

The erroneous cases above were provisionally excluded from the dictionary (see Section 7). Validation by a domain specialist was considered necessary.

Ambiguities

Lexical ambiguities are a significant problem in Natural Language Processing generally, and our dataset was no exception. A word in the UMLS can be assigned more than one CUI, with each CUI having either a different semantic type or the same semantic type as the other CUIs. Take, for example, the CUI C1956346: Coronary
Disease (Coronary Artery Disease) [Disease or Syndrome] and C0010068: Coronary Disease (Coronary heart disease) [Disease or Syndrome]. As another example, where more than one CUI have been assigned to different morphological types of a word, consider C0027740: Nerves (Nerve) [Tissue] and C1280541 Nerve (Entire nerve) [Body Part, Organ, or Organ Component].

Sometimes named entities themselves were ambiguous. For example, in the phrase “[…] 22 out of the 265 patients (8.3%) treated with Abilify Maintena had symptoms coming back within 26 weeks, compared with 21 out of 266 (7.9%) patients treated with aripiprazole taken by mouth”, back was mapped as the named entity C0004600: BACK (Back) [Body Location or Region] because the particle verb come back was not properly recognized.

Another example of a named entity being ambiguous is in the sentence “It is available as prolonged-released capsules containing tacrolimus”, capsule, the small container filled with medicine, was recognized as a Cell Component: C1325531: capsule (Microbial anatomical capsule structure) [Cell Component].

The process of English word formation from the Greek language, especially in sciences, also raised ambiguities, e.g. in Greek, the word θρόμβος is used for both Blood clot [body substance] (Anatomy) and Thrombus [pathologic function] (Disorder).

Variants

Due to diglossia, different word forms in Katharevousa and Demotic Greek can co-exist in the corpus. We found examples that had to be taken into account, e.g. κοιλία and κοιλιά: σε περίπτωση χειρουργικής επέμβασης κοιλίας in abdominal surgery διοχετεύεται υγρό στην κοιλιά fluid is pumped into the abdomen

A variant of this problem can also be observed in English. For example, tummy or belly (non-specialized vocabulary) are often used to refer to abdomen.

To Cetrotide χορηγείται με ένεση κάτω από το δέρμα της κάτω κοιλιακής χώρας. Cetrotide is given by injection under the skin of the lower abdomen (tummy)

διάταση κοιλίας (προσμένη κοιλιά) abdominal distension (swollen tummy)

κοιλιακό άλγος […] πόνος στην περιοχή της κοιλιάς abdominal pain […] pain in the belly area

Spelling variants in Greek, such as διαβητικό προκώμα/διαβητικό προ-κώμα (diabetic pre-coma or diabetic precoma), had also to be taken under consideration.

7 Ambiguities having an equal score were resolved via the option Word Sense Disambiguation provided by MetaMap.
7. Corpus Processing

After the process of recognition, Greek biomedical named entities were inserted in a dictionary in order to obtain all relevant inflected forms.

Textual information was processed by a multilingual platform, Unitex (available at http://www-igm.univ-mlv.fr/~unitex). This multi-system software allows natural language processing and provides a great variety of tools to computational linguists. Unitex has no difficulty in working with Greek and it can handle all language characters. We based our morphological analysis on the application of electronic dictionaries and grammars to our corpus. The general language dictionaries of Greek simple and compound words have been constructed by the Computational Linguistics Research Team of the Laboratory of Translation and Natural Language Processing (Kyriacopoulou, Tsaknaki & Tziafa 2009).

We also developed complex machine-readable grammars, following the formalism of Recursive Transition Networks (RTNs), which are often used to represent linguistic data (Gross 1993; Roche & Schabès 1997). These transition networks contain semantic constraints imposed by specialized language used in medicine while at the same time they describe a great variety of syntactic structures of common language. Thus, these Finite State Transducers (FSTs) are not locally applied to the corpus but are successively applied in order of priority so as to achieve the best global results. Their application was very helpful in the exploration of semantic relations as described in the following section.

8. Semantic Relations

Our research also included the study of semantic relations. Based on the links between Semantic Types – provided by the Semantic Network – and further observations we realized in the sample corpus, we manually recorded observed linguistic patterns (Lee 2004; Abacha & Zweigenbaum 2011). In doing so, we attempted to detect the semantic relationships between ANATOMY and DISORDERS. Further
study on an extended corpus is needed. The following categories embedded in the corpus have been observed:

- **A**djective [**ANATOMY**] + **N**oun [**DISORDERS**]
  - ηπατική δυσλειτουργία
  - reduced liver function

- **N**oun [**DISORDERS**] + (DET erminer:Genitive) / (μέσα (in)) PREP osition DET erminer + **N**oun [**ANATOMY**]
  - σπάσιμο των οστών
  - broken bones
  - πόνος στο στομάχι
  - stomach ache

- **N**oun [**DISORDERS**] (+ RELativePROnoun) (+ negation) + Verb (+ PREPosition) (DETerminer) + **N**oun [**ANATOMY**]
  - καρκίνος η οποία καθιστά τα οστά εύθραυστα
  - cancer that does not affect the bone marrow

- **N**oun [**DISORDERS**] (+ RELativePROnoun) + καθιστώ (make) + DETerminer + **N**oun [**ANATOMY**] + A**djective** [**DISORDERS**]
  - ασθένεια η οποία καθιστά τα οστά εύθραυστα
  - disease that makes bones fragile

- **N**oun [**ANATOMY**] (+ RELativePROnoun) + Vverb (+ PREPosition) (DETerminer) + **N**oun [**DISORDERS**]
  - νευρικών κυττάρων που έχουν υποστεί βλάβη
  - nerve cells that have been damaged

- **N**oun [**ANATOMY**] + γίνομαι (become) + A**djective** [**DISORDERS**]
  - οστά γίνονται λεπτά και εύθραυστα
  - bones become thin and fragile

An important issue we had to deal with was the expected problem of the complex syntactic structure of sentences which were partly resolved e.g. with respect to co-ordination (Foufi, Ioannidou & Tsaknaki 2013). Nevertheless, a number of cases remain to be studied. In the following example, ιστός (tissue) (**N**oun [**ANATOMY**]) is related syntactically to βλάβη (damage) (**N**oun [**DISORDER**]) and they are both complements of προστασία (protection):[8]

προστασία των ιστών από βλάβες
to protect the tissue from damage

Anaphora also extends the list of problems and is expected to be studied in the future:

στα νεφρά και στους οφθαλμούς, και τα καταστρέφει
in the kidneys and the eyes, damaging them

---
[ 721 ]
9. Results

We assessed the results concerning the recognition of named entities using the standard metrics of recall and precision values and the F1 score. In the context of the present study, precision can be defined as the ratio of relevant named entities over the total number of the ones (relevant and irrelevant) retrieved by the corpus processing system. Recall is the ratio of relevant named entities retrieved by the corpus processing system over the total number of relevant named entities which exist in the corpus. Given the size of the whole corpus, we evaluated our method in 100 EPAR, different from the ones already used as a sample corpus.

Our system achieved the following precision and recall rates:

<table>
<thead>
<tr>
<th></th>
<th>Anatomy</th>
<th>Disorders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precision</td>
<td>87%</td>
<td>92%</td>
</tr>
<tr>
<td>Recall</td>
<td>78%</td>
<td>67%</td>
</tr>
<tr>
<td>F1 score</td>
<td>82%</td>
<td>77.5%</td>
</tr>
</tbody>
</table>

Precision is high, as expected, since our method was strictly based on linguistic data. Nevertheless, our resources will be optimized in order to get better results for recall and keep a balance. A large number of false positives were detected to ὅργανος, which designates both body and organization. It was erroneously recognized as a body part (ANATOMY) whereas in the text, the word corresponded to organization.9 Therefore, the performance of our method has been degraded significantly. Based on our observations such occurrences may be large in number but concern only a small number of different words. This means that in this specific domain, there are a few words, used in a great frequency, that are ambiguous. These words can be collected, at a later stage in our process, and ambiguities can be resolved by the use of context (FSTs). As for the recall, false negatives were not recognized because they were not inserted in the FSTs as they should have been. Spelling or typing mistakes also influenced the results. In addition, synonyms of some biomedical named entities were not always recognized. For instance, airways is usually expressed in our corpus with αεραγωγός. However, in some cases αεροφόρα ὁδός was used instead and, thus, has not been recognized by our method. An exhaustive listing of all named entities detected in our aligned corpus will greatly improve the results.

On the contrary, if two or more English synonyms had the same equivalent in Greek, the latter was always recognized even if one of the English synonyms was not captured by MetaMap. For instance, θεραπεία was used in the Greek language version for treatment and therapy. From these two named entities, only therapy was recognized.

9 See Section 6.
Finally, parsing problems could not always be resolved. For instance, parentheses are used a lot in EPAR between the components of named entities, e.g.:

σοβαρή σύσφιγξη (στένωση) των αιμοφόρων αγγείων των πνευμόνων
severe constriction (narrowing) of the blood vessels of the lungs

10. Future work

In this article, we presented a method for recognizing biomedical named entities and extracting semantic relations between them. This research is currently in a developing state. Our corpus in its entirety will serve as an operational tool to attain generalized observations. Future work should cover, among other issues:

- The enrichment of the finite state transducers (FSTs) in order to enhance the precision rates (e.g. elimination of ambiguities).
- The addition of new entries in the Greek biomedical named entities dictionary, by processing a larger corpus.
- The collection of synonymous named entities (e.g. heart failure, cardiac failure, myocardiac failure).
- The use of syntax rules in combination with existing finite state transducers (FSTs).
- The addition of more semantic relations between [ANATOMY] and [DISORDERS], by studying a larger corpus.
- An intracategorical analysis. During our research we noticed other relations inside each Semantic Group we dealt with, e.g. βλάβη [DISORDERS] | στα νεύρα [ANATOMY] (damage to the nerves in the extremities), consequently a thorough study of semantic relations inside [ANATOMY] or [DISORDERS] should be carried out. Additionally, adjectives denoting intensity, duration, degree, e.g. φλεγμονή μεγάλης διάρκειας στους αεραγωγούς των πνευμόνων (long-lasting inflammation of the airways in the lungs) or βαριάς μορφής πόνος στον θώρακα (a severe type of chest pain) play an important role.
- The further study of other semantic relations, e.g.: Η αιμολυτική αναιμία [Disease or Syndrome] ενδέχεται να βλάψει την καρδιακή λειτουργία [Organ or tissue function] του ασθενούς; The haemolytic anaemia may affect a patient’s heart function.

Improvement of each of the above areas should contribute to the total system performance.

References


Chernyshevich, M. & V. Stankevitch. 2015. IHS-RD-BELARUS: Clinical Named Entities Iden-
O. TSAKNAKI & K. IOANNIDOU


Tran, L., G. Divita, M. Carter, J. Judd, M. Samore & A. Gundlapalli. 2015. Exploiting the UMLS Metathesaurus for extracting and categorizing concepts representing signs and
NAMED ENTITY RECOGNITION AND SEMANTIC RELATION EXTRACTION

symptoms to anatomically related organ systems. *Journal of Biomedical Informatics* 58, 29–27.


Keywords: biomedical named entities, recognition, semantic relations, Modern Greek