

## **THE APPLICATION OF CONCEPTNET FOR KNOWLEDGE MANAGEMENT OF TEXTS IN THE POLISH AND ENGLISH LANGUAGES**

WALDEMAR KARWOWSKI, MARIAN RUSEK

*Department of Informatics, Warsaw University of Life Sciences (SGGW)*

In the paper, the possibility of ConceptNet application for knowledge management of texts in the Polish and English language is discussed. ConceptNet is a freely-available semantic network, designed to help computers understand the meanings of words that people use, it contains concepts from many languages. The semantics of ConceptNet is presented in the context of the semantics of the EuroVoc which is based on the SKOS standard. Indexing tool, created by authors, based on ConceptNet is briefly described. Tests of texts indexing with the use of papers in Polish and English language were performed. Results are analyzed and evaluated. At the end, some conclusions about the quality of results are formulated.

Keywords: Knowledge management, Semantic network, Text indexing

### **1. Introduction**

Nowadays people have the enormous amount of information contained in the various types of documents. Excess information makes it difficult to find the content people looking for. Fortunately, summaries and keywords can help during a selection of the most interesting texts. Today documents are stored primarily in digital form in huge computer archives. To manage archives of documents, software tools that perform tasks such as automatic indexing are very useful. Especially in the Internet, search engines use specific algorithms for texts indexing to help find interesting information. Such tools like Google or Bing work very well. How-

ever search engines are not always the best solution, although using the advanced search option. It is often important that the theme and vocabulary should be precisely defined and then we need tools based on the dictionaries built on precisely specified standards. Additionally, it is important to search for semantically similar information. To ensure this, we need thesauruses or ontologies that contain semantic relationships between words. Moreover, in today's organizations information must be available at the same time in different languages. This is the case in international corporations and also in UN agencies or the EU institutions. It means that tools must enable simultaneous search in multiple languages.

The main aim of the work was to evaluate the usefulness of ConceptNet [1], a freely-available semantic network, for multilingual, Polish and English, indexing. Especially we concentrated on the quality of semantic relations included in ConceptNet and use them to determine the keywords. An essential part of the task was to perform a short review of standards, especially SKOS and EuroVoc dictionary, which support semantic relationships and multilingualism and compare them with ConceptNet. An additional goal was an implementation of the simple indexation system based on ConceptNet to test its capabilities in practice.

The rest of this paper is organized as follows: in Sect. 2 the standards of thesauri together with analysis of semantic possibilities are discussed. In Sect. 3 ConceptNet capabilities are presented together with indexation application based on them. We finish with test results presenting, evaluation of usefulness of ConceptNet and brief final remarks in Sect. 4.

## 2. Multilingual thesauri standards

For management of knowledge included in texts, we need dictionaries with a lot of information about the word semantic relationships. The most interesting type of dictionary is a thesaurus. Two standards for thesauri were defined by ISO. The first, ISO 2788, is titled "Guidelines for the establishment and development of monolingual thesauri"; it was published in 1974 and revised in 1986. The second, ISO 5964, was published in 1985 and titled "Guidelines for the establishment and development of multilingual thesauri"; it is connected with the ISO 2788. The two mentioned standards were joined and extended as ISO 25964 "Information and documentation - Thesauri and interoperability with other vocabularies" [2]. This standard contains two parts: "Thesauri for information retrieval" published in 2011, and "Interoperability with other vocabularies" published in 2013. Semantic in ISO 25964 is based on *Thesaurus Concept*, this is something that exists in the mind as abstract entities independent of terms used to express them; word or phrase used to label a concept is called *Thesaurus Term*. Every concept has the *Preferred Term* and may have *Simple Non Preferred Term*. Between synonyms and near-synonyms we have equivalence relation - *USE*. It suggests the use of a different term (pointed

by *USE*). Relation *UF* - use for – means reciprocal to *USE*. There is additionally defined *Compound Equivalence* when a phrase corresponds to a particular term. Hierarchical relationship (between broader and narrower concepts) can be defined by *Broader Term* (BT) and *Narrower Term* (NT). Each concept can have a pointer linking it to the concept at the top of any hierarchy in which it occurs. Associative relationship *Related Term* (RT) is defined between concepts that are closely related in some non-hierarchical way and represents a preferred term for the concept that has an association with the one. Possibility of adding attributes and labels make semantic stronger. Moreover, groups of sibling concepts, which have a common parent concept, may be organized into arrays introduced by node labels. We have to note that the old standards are still important in the sense that multiple implementations of thesauri are based on old versions. Relations in older standards are analogous to 25964 but everything is based on terms (i.e. term are not clear distinguished from concept).

W3C recommendation Simple Knowledge Organization System (SKOS) [3] is a part of Semantic Web standards. SKOS is an XML dialect built upon RDF and RDFS, other parts of the Semantic Web family of standards. Additionally SKOS was extended by extension for labels (SKOS-XL), which was defined to provide additional support for describing and linking lexical entities [4]. SKOS was designed with a big influence of described above ISO-2788/5964 standards. Generally, it is a big correspondence between them and it is easy to convert one to the other, details about this correspondence are described in Appendix of [5] and [6, 7]. The main difference is that ISO standards focus on how to build and manage a thesaurus, but SKOS focuses on how to publish a vocabulary in a format easy to process by computer and compliant with the Semantic Web standards. The SKOS data model is concept-based. SKOS supports rich semantics for classes (Concept, Collection, Concept Scheme), object properties (has broader, has broader match, has broader transitive, has close match, has exact match, has member, has member list, has narrower, has narrower match, has narrower transitive, has related, has related match, has top concept, is in mapping relation with, is in semantic relation with), and annotation properties (preferred label, alternative label, note etc.). More information about other similar standards was presented in [8].

Multilingual thesaurus EuroVoc [9] is an example of ISO standards implementation and at the same time SKOS and SKOS-XL implementation. It is multidisciplinary and covers all issues relating to the activities of the European Parliament and contains terms in 23 EU languages (Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Italian, Latvian, Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovenian, Spanish and Swedish), 24-th version Gaelge i.e. Irish is under translation. Moreover EuroVoc has versions in three languages of UE candidate countries: Macedonian (македонски - mk), Albanian (shqip - sq) and Serbian (српски - sr). The dic-

tionary is intended for use during simultaneous translation, preparation of documents and to provide a controlled vocabulary for the information management software tools. It contains over 7000 terms grouped into domains and subdomains. SKOS and SKOS-XL implementation (file in RDF format) The EuroVoc is available in ISO standards through web page additionally user can download it as the set of pdf documents. Moreover, digital version in SKOS/RDF format is possible to download. The EuroVoc is based on dedicated ontology defined in SKOS and SKOS-XL standards. For programmers, set of SKOS/RDF files and supplemental XML files are provided together with simple ReST service with limited functionality.

Through the web interface, we can obtain semantic relations of selected term (concept). In the Figure 1 result for “public administration” concept is presented in the Polish language on the left side and English on the right side. A similar result can be obtained from pdf files.

<p><b>administracja publiczna</b>  <b>04 ŻYCIE POLITYCZNE</b>            MT 0436 władza wykonawcza i służba publiczna            NT1 administracja centralna                NT2 ministerstwo            NT1 administracja elektroniczna            NT1 administracja lokalna            NT1 administracja regionalna            NT1 analiza polityki [4.8]            NT1 instytucja publiczna            NT1 kształtowanie polityki            NT1 nadzór administracyjny            NT1 partnerstwo publiczno-prywatne            NT1 polityka władz publicznych            NT1 przedstawiciele samorządów lokalnych i regionalnych            NT1 siedziba organu            NT1 służba publiczna            NT1 stosunki administracja-obywatel                NT2 formalności administracyjne                NT2 jawność administracyjna                NT2 rzecznik praw obywatelskich                NT2 zażalenie administracyjne            RT budynek publiczny [ 2846 ] nauki administracyjne [ 3611 ] organy państwowe [ 0406 ] władza dyskrecyjna [ 0406 ]  <b>URI</b> <a href="http://eurovoc.europa.eu/77">http://eurovoc.europa.eu/77</a>  <b>Has Exact Match</b>  <i>Public administration (AGROVOC) administration (AGROVOC) Public administration (ECLAS) administration (GEMET) Public administration (STW Thesaurus for Economics) public administration (Unbis Thesaurus)</i></p>	<p><b>public administration</b>            UF <i>general government</i>  <b>04 POLITICS</b>            MT 0436 executive power and public service            NT1 administration headquarters            NT1 administrative supervision            NT1 central government                NT2 ministry            NT1 citizen-authority relations                NT2 administrative formalities                NT2 administrative transparency                NT2 appeal to an administrative authority                NT2 mediator            NT1 electronic government            NT1 local government            NT1 policy analysis [4.8]            NT1 policymaking            NT1 public institution            NT1 public policy            NT1 public-private partnership            NT1 public service            NT1 regional government            NT1 representative of local or regional authority            RT administrative science [ 3611 ] discretionary power [ 0406 ] public authorities [ 0406 ] public building [ 2846 ]  <b>URI</b> <a href="http://eurovoc.europa.eu/77">http://eurovoc.europa.eu/77</a>  <b>Has Exact Match</b>  <i>Public administration (AGROVOC) administration (AGROVOC) Public administration (ECLAS) administration (GEMET) Public administration (STW Thesaurus for Economics) public administration (Unbis Thesaurus)</i></p>
---	--

**Figure 1.** Concept of “public administration”  
 Source: EuroVoc web page [10]

Terms are described according to ISO standard, NT1 means narrower term one level down, NT2 two levels down. UF denotes that public administration is preferred term for “general government”. Contrary “general government” has USE relation with “public administration”. We have to note that every concept has its own identifier; “public administration” has 77. Related concepts similarly have unique identifiers “administrative science” 3611, “discretionary power” – 0406, “public authorities” – 0406, “public building” - 2846. Additionally, we have some extensions: domain – 04 POLITICS, and subdomain - MT 0436 executive power and public service. Moreover, analogous terms in other dictionaries are presented in HasExactMatch section. We can easily obtain versions in other languages with the web interface or from files with parallel concepts listings. We can see in the Polish language we have no synonyms.

```

<rdf:Description rdf:about="http://eurovoc.europa.eu/77">
  ...
  <dc:identifier>77</dc:identifier>
  ...
  <skos:prefLabel xml:lang="nl">overheidsadministratie</skos:prefLabel>
  <skos:prefLabel xml:lang="en">public administration</skos:prefLabel>
  ...
  <skos:altLabel xml:lang="en">general government</skos:altLabel>
  <skos:altLabel xml:lang="de">Behörde</skos:altLabel>
  ...
  <skos:prefLabel xml:lang="pl">administracja publiczna</skos:prefLabel>
  <skos:altLabel xml:lang="cs">orgány státní správy</skos:altLabel>
  ...
  <skosxl:prefLabel rdf:resource="http://eurovoc.europa.eu/209582"/>
  ...
  <skos:narrower rdf:resource="http://eurovoc.europa.eu/6894"/>
  ...
  <skos:prefLabel xml:lang="fi">julkinen hallinto</skos:prefLabel>
  <skosxl:prefLabel rdf:resource="http://eurovoc.europa.eu/328377"/>
  ...
</rdf:Description>

```

**Figure 2.** Concept of “public administration”

*Source:* EuroVoc SKOS/RDF file [11]

In the Fig. 2 fragment of SKOS/RDF file for “public administration” concept is presented. As we can see all information is included in <rdf:Description rdf:about="http://eurovoc.europa.eu/77"> element with identifier 77. There are preference label for Polish term <skos:prefLabel xml:lang="pl"> administracja publiczna</skos:prefLabel> and for English term <skos:prefLabel xml:lang="en">public administration</skos:prefLabel>. We have to note that terms are defined separately in SKOS document and have its own identifiers (Fig. 3), in the Fig. 2 we have listed English preference label second time as <skosxl:prefLabel rdf:resource="http://eurovoc.europa.eu/209582"/>.

Preference label for term in Polish “administracja elektroniczna” is listed second time too: `<skosxl:prefLabel rdf:resource="http://eurovoc.europa.eu/328377"/>` (Fig. 3). Narrower or broader terms are pointed by their identifier for “electronic government” we have `<skos:narrower rdf:resource="http://eurovoc.europa.eu/6894"/>`.

```

<rdf:Description rdf:about=http://eurovoc.europa.eu/209582
  <rdf:type rdf:resource="http://www.w3.org/2008/05/skos-xl#Label"/
  <owl:versionInfo rdf:datatype="http://www.w3.org/2001/XMLSchema#string">n/a</owl:versionInfo>
  <skosxl:literalForm xml:lang="en">public administration</skosxl:literalForm>
  <dct:type rdf:resource="http://publications.europa.eu/resource/authority/label-type/STANDARDLABEL"/>
</rdf:Description>
...
<rdf:Description rdf:about="http://eurovoc.europa.eu/328377">
  <rdf:type rdf:resource="http://www.w3.org/2008/05/skos-xl#Label"/>
  <owl:versionInfo rdf:datatype="http://www.w3.org/2001/XMLSchema#string">n/a</owl:versionInfo>
  <skosxl:literalForm xml:lang="pl">administracja publiczna</skosxl:literalForm>
  <dct:type rdf:resource="http://publications.europa.eu/resource/authority/label-type/STANDARDLABEL"/>
</rdf:Description>

```

**Figure 3.** Terms “public administration” and “administracja publiczna”  
*Source:* EuroVoc SKOS/RDF file [11]

EuroVoc like ConceptNet is dedicated to the wide range of issues. EuroVoc is not only one example of SKOS implementation. There are many dictionaries dedicated to specific topics. Very well-known is AGROVOC [12], huge multilingual thesaurus in the agriculture domain. EuroVoc has fewer extensions to SKOS format than AGROVOC and allows for a simpler illustration of semantic relationships. More information about other multilingual dictionaries was presented in [8]. Applications of AGROVOC to indexing agricultural texts in Polish and English language were discussed by us in [8]. More examples of indexing tools are presented in [8] and [16].

### 3. ConceptNet

ConceptNet is a freely-available semantic network, designed to help computers understand the meanings of words that people use [1]. It started from the crowdsourcing project Open Mind Common Sense, which was launched in 1999 at the MIT Media Lab. At the beginning the facts were collected from people who came to the Open Mind Common Sense site. Today ConceptNet is a huge graph of general knowledge representing words and phrases and the relationships between them. The data is taken from a variety of resources. Big collection of facts is acquired from Open Mind Common Sense. Much information is extracted from parsing Wiktionary, in multiple languages. Wiktionary gives information about synonyms, antonyms, translations of concepts into hundreds of languages, and multiple

labeled word senses for many words. The next source of knowledge is games with a purpose (such as Verbosity and nadya.jp) designed to collect common knowledge. These games give knowledge about people’s intuitive word associations. Expert-created dictionary-style knowledge comes from WordNet, Open Multilingual WordNet and JMDict (Japanese-multilingual dictionary). Some knowledge comes from OpenCyc, a hierarchy of hypernyms provided by Cyc a system that represents common sense knowledge in predicate logic. Moreover, ConceptNet is connected to a subset of DBpedia, which extracts knowledge from Wikipedia articles. With the combination of these sources, ConceptNet contains over 21 million edges and over 8 million nodes. Its English vocabulary contains approximately 1,500,000 nodes, and there are 83 languages in which it contains at least 10,000 nodes. The largest source of input for ConceptNet is Wiktionary, which provides 18.1 million edges and is mostly responsible for its large multilingual vocabulary [13].

Terms and phrases are nodes in ConceptNet graph. Generally nodes (terms) are identified by URLs, for example, /c/en/vegetable/, and language code is a part of an identifier. Nodes are connected with edges. An edge, or assertion, is a unit of knowledge which describes a particular relation between two nodes - natural-language terms. Every assertion has a relation: start node and end node; it means that generally assertions are directed. Rich semantic is supported by 36 core relations. There are defined 7 symmetric relations: Antonym, DistinctFrom, EtymologicallyRelatedTo, LocatedNear, RelatedTo, SimilarTo, and Synonym. The directionality of symmetrical edges is not important; start node can be exchanged with end node. Moreover there are defined twenty nine asymmetric relations: AtLocation, CapableOf, Causes, CausesDesire, CreatedBy, DefinedAs, DerivedFrom, Desires, Entails, ExternalURL, FormOf, HasA, HasContext, HasSubevent, HasFirstSubevent, HasLastSubevent, HasPrerequisite, HasProperty, InstanceOf, IsA, MadeOf, MannerOf, MotivatedByGoal, ObstructedBy, PartOf, ReceivesAction, SenseOf, SymbolOf, and UsedFor. For example, IsA relation describes hierarchical relation and means that start node is a subtype or a specific instance of end node. We have to note that there is separate relation InstanceOf which means that start node is an example of end node. Some asymmetric relations are opposite to each other, for example PartOf is reverse to HasA. Synonym means that start node and end node have very similar meanings. It should be emphasized that Synonym may be term translation in a different language. Very important is FormOf relation it describes that start node is an inflected form of end node; end node is the root word of start node. We illustrate above remarks on examples, starting from “public administration”.

In the Fig. 4 a fragment of JSON file for “administracja publiczna” is presented. The file is relatively small; node /c/pl/administracja\_publiczna/ has only two edges. The first edge is relation r/ExternalURL/ which connects it with node

/http://pl.dbpedia.org/resource/Administracja\_publiczna/ (for short in Fig. 4 only id is presented). The second edge is relation r/Synonym/ and connects it with English translation node /c/en/public\_administration/. This edge is presented in its entirety in the Fig. 4 to show a full edge description. The edge details: start node, end node and sources of information are described.

```

{
  "@context": [ "http://api.conceptnet.io/ld/conceptnet5.6/context.ld.json" ],
  "@id": "/c/pl/administracja_publiczna",
  "edges": [
    { "@id": "/a[/r/ExternalURL/,/c/pl/administracja_publiczna/,http://pl.dbpedia.org/resource/Administracja_publiczna/]",... },
    { "@id": "/a[/r/Synonym/,/c/pl/administracja_publiczna/,/c/en/public_administration/]",
      "@type": "Edge",
      "dataset": "/d/dbpedia/en",
      "end": {
        "@id": "/c/en/public_administration",
        "@type": "Node",
        "label": "public administration",
        "language": "en",
        "term": "/c/en/public_administration"
      },
      "license": "cc:by-sa/4.0",
      "rel": {
        "@id": "/r/Synonym",
        "@type": "Relation",
        "label": "Synonym"
      },
      "sources": [
        {
          "@id": "/s/resource/dbpedia/2015/en",
          "@type": "Source",
          "contributor": "/s/resource/dbpedia/2015/en"
        }
      ],
      "start": {
        "@id": "/c/pl/administracja_publiczna",
        "@type": "Node",
        "label": "administracja publiczna",
        "language": "pl",
        "term": "/c/pl/administracja_publiczna"
      },
      "surfaceText": null,
      "weight": 0.5
    }
  ]
}

```

**Figure 4.** Term “administracja publiczna”.  
*Source:* ConceptNet API [14]

In the Fig. 5 fragments of JSON file for “public administration” are presented. There are more than 70 edges. Most of them are edges with Synonym relation -



translation to other languages. It is interesting that some Synonyms, even it is symmetric relation, are present as two edges, for example for French translation we have `/a/[r/Synonym/,c/fr/administration_publique/, /c/en/public_administration/]` and `/a/[r/Synonym/,c/en/public_administration/n/, /c/fr/administration_publique/]`. For most languages, among them Polish, we have only one edge start node is in particular language, but end node is in English. For the Polish language we have `/a/[r/Synonym/,c/pl/administracja_publiczna/, /c/en/public_administration/]`. For the Japanese language we have three edges two edges `/a/[r/Synonym/,c/ja/行政学/n/,c/en/public_administration/]`, and one edge `/a/[r/Synonym/,c/en/public_administration/n/,c/ja/行政/]`. It is connected with source of information. It is Japanese-multilingual dictionary (jmdict/1.07), DBPedia (dbpedia/2015/en) and Wiktionary (wiktionary/en). For some languages, like French, there are two sources DBPedia (dbpedia/2015/en) and Wiktionary (wiktionary/en). For Polish, there is only one source DBPedia (dbpedia/2015/en).

```

{
  "@context": [ "http://api.conceptnet.io/ld/conceptnet5.6/context.ld.json" ],
  "@id": "/c/en/public_administration",
  "edges": [
    {"@id": "/a/[r/Synonym/,c/ja/行政学/n/,c/en/public_administration/]", ... },
    {"@id": "/a/[r/ExternalURL/,c/en/public_administration/,http://dbpedia.org/resource/Public_administration/]",...
    },
    {"@id": "/a/[r/ExternalURL/,c/en/public_administration/,http://wikidata.dbpedia.org/resource/Q31728/]", ... },
    {"@id": "/a/[r/RelatedTo/,c/en/public_administration/n/,c/en/administration/]", ... },
    ...
    {"@id": "/a/[r/Synonym/,c/en/public_administration/n/,c/de/öffentliche_verwaltung/]",... },
    {"@id": "/a/[r/Synonym/,c/en/public_administration/n/,c/fr/administration_publique/]",... },
    {"@id": "/a/[r/Synonym/,c/en/public_administration/n/,c/ja/行政/]",... },
    ...
    {"@id": "/a/[r/Synonym/,c/en/public_administration/n/,c/zh/公共行政/]",... },
    {"@id": "/a/[r/Synonym/,c/ar/إدارة_علمة/,c/en/public_administration/]",... },
    ...
    {"@id": "/a/[r/Synonym/,c/fr/administration_publique/,c/en/public_administration/]",... },
    ...
    {"@id": "/a/[r/Synonym/,c/ja/行政学/,c/en/public_administration/]",... },
    ...
    {"@id": "/a/[r/Synonym/,c/pl/administracja_publiczna/,c/en/public_administration/]",... },
    ...
    {"@id": "/a/[r/Synonym/,c/zh/公共行政学/,c/en/public_administration/]",... },
    {"@id": "/a/[r/ExternalURL/,c/en/public_administration/,http://en.wiktionary.org/wiki/public_administration/]",
    ... }
  ],
  ...
}

```

**Figure 5.** Term “public\_administration”  
*Source:* ConceptNet API [14]

From edges with relation FormOf we can obtain inflected forms. To illustrate this we can show edges for the term “administracja”. In the Fig. 6 fragment of JSON file for “administracja” is presented. We have inflected forms: *administracjach*, *administracją*, *administracjami*, *administracje*, *administrację*, *administracji*, *administracjo*, *administracjom*, *administracyj*. All forms (except the last) appear two times, the edges differ only by the noun designation (/n) and additionally we have edges with the same forms and RelatedTo relation. Sources of inflected forms are French and German Wiktionary. There are two additional edges with RelatedTo: */a[/r/RelatedTo/,c/pl/administracja/n/,c/fr/administration/]* and */a[/r/RelatedTo/,c/pl/administracja/n/,c/en/administration/]*. Moreover there are three edges with Polish Synonyms: *rząd*, *zarząd*, *zarządzać* and */a[/r/Synonym/,c/en/administration/n/,c/pl/administracja/]*.

```

{
  "@context": [ "http://api.conceptnet.io/ld/conceptnet5.6/context.ld.json" ],
  "@id": "/c/pl/administracja",
  "edges": [
    { "@id": "/a[/r/DerivedFrom/,c/pl/administracja/,c/pl/acja/]",... },
    { "@id": "/a[/r/DerivedFrom/,c/pl/administracyjny/,c/pl/administracja/n/]",... },
    { "@id": "/a[/r/FormOf/,c/pl/administracjach/,c/pl/administracja/n/]",... },
    { "@id": "/a[/r/FormOf/,c/pl/administracjach/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/FormOf/,c/pl/administracją/,c/pl/administracja/n/]",... },
    { "@id": "/a[/r/FormOf/,c/pl/administracją/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/FormOf/,c/pl/administracjami/,c/pl/administracja/n/]",... },
    { "@id": "/a[/r/FormOf/,c/pl/administracjami/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/FormOf/,c/pl/administracjom/,c/pl/administracja/n/]",... },
    ...
    { "@id": "/a[/r/FormOf/,c/pl/administracjom/,c/pl/administracja/n/]",... },
    { "@id": "/a[/r/FormOf/,c/pl/administracjom/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/FormOf/,c/pl/administracjo/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/FormOf/,c/pl/administracyj/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/IsA/,c/pl/adm/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/RelatedTo/,c/pl/administracjach/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/RelatedTo/,c/pl/administracjami/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/RelatedTo/,c/pl/administracja/n/,c/en/administration/]",... },
    { "@id": "/a[/r/RelatedTo/,c/pl/administracja/n/,c/fr/administration/]",... },
    { "@id": "/a[/r/RelatedTo/,c/pl/administracja/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/RelatedTo/,c/pl/administrację/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/RelatedTo/,c/pl/administrację/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/RelatedTo/,c/pl/administracji/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/RelatedTo/,c/pl/administracjom/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/RelatedTo/,c/pl/administracjo/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/RelatedTo/,c/pl/adm/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/Synonym/,c/en/administration/n/,c/pl/administracja/]",... },
    { "@id": "/a[/r/Synonym/,c/pl/administracja/n/,c/pl/rząd/]",... },
    { "@id": "/a[/r/Synonym/,c/pl/administracja/n/,c/pl/zarząd/]",... },
    { "@id": "/a[/r/Synonym/,c/pl/administracja/n/,c/pl/zarządzać/]",... },
    { "@id": "/a[/r/ExternalURL/,c/pl/administracja/,http://en.wiktionary.org/wiki/administracja/]",... },
    { "@id": "/a[/r/ExternalURL/,c/pl/administracja/,http://fr.wiktionary.org/wiki/administracja/]",... }
  ]
}

```

**Figure 6.** Term “administracja”

Source: ConceptNet API [14]

Summarizing, in ConceptNet we have much richer semantics than in EuroVoc, however, it is rather chaotic. Synonym relation is ambiguous; translation can be deduced only from the language code. There are no preferred terms. Relation IsA is analogous to broader term, but not the analogy to narrower term. The reason is that to describe narrower term ConceptNet uses the same relation IsA. For example in the description of the term cattle we have ingoing edge /a/[r/IsA/,c/en/cow/n/,c/en/cattle/n/] and in the description of the term cow we have exactly the same but outgoing edge /a/[r/IsA/,c/en/cow/n/,c/en/cattle/n/].

#### **4. Tests results, conclusions and future work**

The indexing tool was prepared in two symmetric versions, to index English text with Polish translation of main words, and to index Polish text with English translation of main words. The indexing algorithm, based on ConceptNet, consists of four steps: removing stopwords; finding the base form of a word by analyzing FormOf edges connected with the word; recognizing the part of speech (noun, verb or adjective); translating founded most frequent words into second language by Synonym edges. For testing English indexation with Polish translation we selected 8 texts published in English language in Agricultural Engineering Journal (vol. 153, 154, 155). We analyzed ten most frequent nouns and five most frequent verbs, because it allows recognizing the links among the texts [15]. Generally indexing in English based on ConceptNet worked quite good, all most important nouns were properly recognized likewise verbs. Finding basic form of word was almost perfect. Automatic translation into Polish was relatively good, even some translation had bad quality. Ten important nouns were translated in 85%, five verbs only in 67% of cases. For testing Polish indexation with translation into English we selected nineteen publications in Polish language from older volumes of the same journal as above (the same papers, particularly abstracts, were used in the analysis performed in [16] to evaluation Polish indexer based on AGROVOC and comparing it to other indexers). The first seven papers are connected with maize cultivation and production; the next twelve generally are connected with potatoes. Indexing in Polish based on ConceptNet worked relatively good, almost all most important nouns were properly recognized. Finding basic form of word was about 90% for nouns and 98% for verb. Unfortunately many verbs were unrecognized because they are absent in ConceptNet. Automatic ten important nouns translation into English was similar to English indexer - 81%. Five important verbs were properly translated in 77%.

To summarize we can conclude that ConceptNet allows finding the basic forms of nouns; however simple algorithm based on ForOf relation is not enough. Situation with parallel translation into second language is much complicated. ConceptNet contains much common knowledge, which in fact hinders the automatic

translation. For example the Polish word *kukurydza* has two English synonyms: corn and maize. On the other hand word corn has four synonyms in Polish: *kukurydza*, *nagniotek*, *odcisk*, *odgniotek*. Similarly the Polish word *krowa* has two English synonyms: cow and bag. Unfortunately relation is not symmetric because at the same time the English word cow has synonyms *klepa*, *krowa*, *prukwa*, *pułło*. Generally the biggest problem is chaotic semantics. Relation FormOf is not consequently used, sometimes it is replaced by RelatedTo. Similar situation is with Synonym relation. In conclusion, it can be pointed that ConceptNet is suitable rather as a helper tool. The base of indexers should be thesauri based on standards such as EuroVoc.

## REFERENCES

- [1] ConceptNet: <http://conceptnet.io/>.
- [2] ISO 25964-1:2011 - *Thesauri and interoperability with other vocabularies - Part 1: Thesauri for information retrieval*
- [3] SKOS Recommendation (18 August 2009) : <http://www.w3.org/TR/skos-reference/>
- [4] SKOS XL: <http://www.w3.org/TR/skos-reference/skos-xl.html>
- [5] SKOS Primer (Note 18 August 2009): <http://www.w3.org/TR/skos-primer/>
- [6] <http://www.niso.org/schemas/iso25964/#schema>
- [7] [http://www.niso.org/apps/group\\_public/download.php/12351/Correspondence%20ISO25964-SKOSXL-MADS-2013-12-11.pdf](http://www.niso.org/apps/group_public/download.php/12351/Correspondence%20ISO25964-SKOSXL-MADS-2013-12-11.pdf)
- [8] Karwowski W., Wrzeciono P., (2016) *Methods of automatic topic mining in publications in agriculture domain*. Inf. Syst. in Management 2016, vol. 6 (3), 192-202.
- [9] EuroVoc thesaurus: <https://publications.europa.eu/en/web/eu-vocabularies/>
- [10] EuroVoc interface: <http://eurovoc.europa.eu/drupal/>
- [11] [http://publications.europa.eu/mdr/resource/thesaurus/eurovoc-20171215-0/skos/eurovoc\\_skos.zip/](http://publications.europa.eu/mdr/resource/thesaurus/eurovoc-20171215-0/skos/eurovoc_skos.zip/)
- [12] AGROVOC thesaurus: <http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus>
- [13] Speer, R., Chin, J., Havasi, C.. (2016) *ConceptNet 5.5: An Open Multilingual Graph of General Knowledge*. Proceedings of 31st AAAI Conf. on Artificial Intelligence.
- [14] ConceptNet API: <http://api.conceptnet.io/>.
- [15] Wrzeciono P., Karwowski W. (2013) *Automatic Indexing and Creating Semantic Networks for Agricultural Science Papers in the Polish Language*, Computer Software and Applications Conference Workshops (COMPSACW), 2013 IEEE 37th Annual, Kyoto, 356-360.
- [16] Karwowski, W., Orłowski A., Rusek, M. (2019) *Applications of multilingual thesauri for the texts indexing in the field of agriculture*. Conference proceedings ACS 2018, Advances in Soft and Hard Computing, AISC, volume 889, Springer, Cham, 185-195