Graph theoretical view on text understanding.

Jure Zupan
National institute of Chemistry, Ljubljana
jure.zupan@ki.si

Keywords: graph theory, cyclic-connected-graph, topological distance, network, text analysis, information content

Received: [6. Oct. 2017] Opisani sistem je orientiran na širše izločanje informacij iz slovenskih besedil kot je samo besedna analiza in označevanje besed. Osnova sta dva programska dela. Prvega sestavlja podatkovna baza (149,000 korenov besed in 3,100 končnic) in označevalec slovničnih parametrov, drugega pa 45,000 samostalnikov in 16,000 glagolov, ki so s skupinami teh besed grupirani po različnih značilnostih v povezan cikličen graf (*connected cyclic graph*). Prvi del izvrši slovnično označevanje besed v tekstu, drugi pa med posameznimi besedami, ki so v grafu razvejane v približno 5,000 hierarhično povezanih skupin besed, izračuna topološke razdalje. Izkazalo se je, da topološko izračunana razdalja med besedami dobro predstavi pomensko razliko/sličnost med njimi. Obe besedni zbirki vsabujeta in obdelujeta pretežni del napogostejših slovenskih besed (cca 160,000 slovenskih besed).

V prispevku so razložene nekatere pasti slovenščine pri obvladovanju več-smiselnosti besedila. Opisana je tudi struktura cikličnega grafa besed (samostalnikov in glagolov) in način izračuna topološke razdalje med besedami. Poudarjena je dvosmernost poti in sprehodov (*paths and walks*) v omenjenem grafu besed in zakaj upoštevanje prepoved spremembe smeri sprehoda (*walk*) prepreči, da besede v grafu niso samo-referenčne (*closed cycle*) in da pri vnosu novih besed ne pride do ustvarjanja neskončnih zank. Dodan je kratek primer analize stavka, ki se konča z matriko topoloških razdalj med posameznimi besedami. Na koncu so diskutirane nekatere možnosti nadaljnjega razvoja hierarhične mreže tako pri prevodu v drug jezik (angleščino) kot tudi v sami slovenski različici.

1 Introduction

The goal of the present work is to increase the amount of information from the text hidden in the so called back-ground information. Back ground information is obtained through a hierarchical ordered word collection (graph) using the topological distance D_{ij} between two arbitrary words i and j in the data-base. D_{ij} is calculated in a graph-theoretical manner ([1]) from the parameters defined by the topology of each word.

The hypothesis of the research is that, if one wants to improve a 'man-machine conversation', a broader pool of crucial information is needed compared to those given by the standard lexical or dictionaries' entries. The question is, is it possible to obtain a reliable numerical measure of the difference between the words? We believe that the crucial information in text understanding are facts most often ignored because they seem to be too simple and/or too obvious. If, for example, within the text a word *string* is encountered the crucial information is not its description as a *slender cord for binding*, or *a line*, what is usually written in dictionaries, but rather the property that a string is a *solid man-made object*. Besides what the word means, the information what the word *does not* mean or represents is nevertheless important as well. For example, the *string* is *not* an animal, a plant, a human, is *not* a liquid, so one can't drink it, etc. For a human being this are obvious facts, but not so for a computer. For any word one should supply as many information about the features the word might have or properties it is identified with. The *string* for example has additional meanings of *a line* and of *one-dimensional graphic concept*. If further on, during the text, the words like *violin*, *horse*, *philosopher*, or *geometry*, is encountered, any numerical evaluation for the distance measure between these words in the form *D*(*string, violin*), *D*(*string, horse*), *D*(*string, philosopher*), *D*(*string, geometry*), etc., would be very beneficial to enhance the information what the word *string* actually means in the context of the particular communication. The more supplemental information is provided about each word, the better the process for building up the understanding of the complex information from a given text could be obtained.

This hypothesis helps to argue that as much the meanings of single words is important, any kind of the *distance measure* between different words and their meanings is important as well. This in turn requires two things; first each word should be represented in a uniform way based on various kind of properties and, second, the words should bi organized in a system that allows definition of a metrics.

2 Hierarchical network of word meanings and features

The solution to the discussed information enhancing problem seems to be the organization of words into network or graph of words linked according to the common features or some other commonly present or absent property(ies). Therefore, the links (branches) between nodes in the graph must contain meaningful information about the relation between the nodes they connect. For example: if one node is labeled *tool* and the other one *object* (*man-made*) the link between them must exhibit the property that the first node (cluster of words) labeled *tools* is a pat of the second node labeled *all man-made object*) and not *vice versa*. At the same time these two nodes (clusters) should occupy positions in the graph much closer to each other than they have to the node labeled *insect*, for example. Either individual words or clusters of words could simultaneously be members of several groups what makes a graph or network to contain cyclic paths in the structure (Figure 1).

Using the above kind of reasoning, a graph of about 25,000 and 87,000 verb and noun meanings (16,000 and 45,000 dictionary lexemes of each word type), respectively, clustered into broader *groups* according to their meanings and properties containing close to 5,000 clusters of words (nodes) was generated ([2]). The closest collection to our data base are the Levine's collection of verb classes ([3]) and Dornseiff's Wortschatz ([4]). There are various internet versions like Visuword® ([5]) and for the Slovenian language the SloWNet ([6]). What the size, i.e. the number of words is concerned, only the Dornseiff's collection has about the same number of verbs (14,000) as our collection. The part of our network containing verbs is based on six main groups (*to exist, to have, to move, to do/to, to think/to create, and to sense/to communicate*) and is already well described in the literature ([7,8]) and is accessible on the web ([9]). The complete structure of verb hierarchy in English language (16,000 verbs and 1000 groups) is given in ([10]). The basic division of nouns has three groups: the *product,* the *nature,* and the *concept*. It can be seen from second part of Table 1. The clusters of verbs and nouns in all levels of hierarchy are of very different sizes (Table 1).

	VERBS (24,626)
Verbs of existing (3,405)	to exist on a specific way (542), verbs to sustain living (1,427), to end existence (299), emission verbs (949), weather verbs (187)
Verbs of having (1,339))	to posses (154), to obtain/take (333), to use possession (288), to negotiate possession (461), to spend possession(102)
Verbs of moving (3,129)	to move (general) (804), to move (specific way) (692), to move (body/parts) (629), to arrive/leave (676), to change movement (206), to do while moving (121)
Verbs of doing (9,663)	to put (2,416), to do (general) (669), to assemble/disassemble (1,340), to change (2,164), to use force/influence (1,322), to do complex tasks (1,751)
Verbs of thinking/creating (1,583)	to create (intellectually) (550), to think (general) (145), to think (specific) (407), to expressing thoughts with symbols (480),
Verbs of communicatin (5,507)	to exchange of information (2,770), verbs of perception (322), to have/response to feelings (883), verbs of social contact (1,531),
	NOUNS (86,799)
nature (31,988)	**nature (non-living)**(3,130) is divided into: nature (general) (10), nature (phenomenon) (521), nature (physical parameter) (151), nature (space) (82), matter (general) (1,359), matter(Earth) (933), matter (outer-space) (84) **nature (living) (28,847)** is divided into: nature (general/broader) (4,218), nature (plant kingdom) (3,111), nature (animal kingdom) (3,431), nature (human) (18,087)
product (19,222)	**product (origin) (552)** divided into: product (origin(human)) (40), product (origin(nature)) (53), product (origin(plant)) (258), product (origin(animal)) (201) **product (human) (18,670)** divided into: product (human(material)) (13,190), product (human(intellectual)) (5,352) product (human(commodity)) (29), creation (general) (5), creation(limitation) (94)
concept (35.589)	**activity (11,645)** is divided into: activity (general) (101), activity (to do something) (3,507), activity (society) (3,045), activity (emotion) (76), activity (sense) (15), activity (existence) (1,068), activity (movement) (1,240), activity (communication) (1,912), activity (possession) (582), activity (mind) (97)

	property (5,943) is divided into: property (action) (323), property (animal) (45), property (broader meaning) (357), property (company) (17), property (device) (90), property (form) (62), property (general) (37), property (human) (2,774), property (mind) (128), property (matter) (267), property (nation) (35), property (number) (13), property (object) (482), property (phenomenon) (42), property (plant) (34), property (procedure) (390), property (religion) (15), property (ruling) (52), property (society) (111), property (sound) (39), property (space) (309), property (status) (159) property (word/speech) (123), group of properties (38), and 8 other groups: **event (1,208), form (3,169), group (1,958), phenomenon (526), procedure (992), result (5,342), space (1,532), state (2,910).**

Table 1. The first two levels of verb (upper part of the table) and noun (lower part of the table) hierarchy according to their common features. In the parentheses the number of words in each group is given. Because individual word may be listed in several groups, the sum of words given in parenthesis is larger than the number of lemmas in the network. The largest groups are printed bold.

On the contrast to the English language, the Slovenian lexical forms of verbs can be well distinguished from those of nouns, however, due to high flexibility of Slovenian declension and conjugation (approximately 20 different forms per noun, verb, adjective, pronoun, and numeral) there are numerous cases where two or even three word types mix. For example the sentence *To je lepo padalo* has two meanings, first *This is a nice parachute* and the second one *It was falling nicely*. In the first case the word *padalo* is a noun (*parachute*) while in the second one is a verb (*to fall*). Following the idea to have all words in one network (graph) both word types are linked in the network on the highest node N_{top}.
It is worthwhile to mention that the same word in different languages has different *sets* of meaning. This is the reason why such a hierarchy cannot be 'blue-printed' from one to another language. The effect of 'lost with translation' is unavoidable: each translated word could be connected to completely different clusters of words than in the original language. For example, the English word *plant* in its botanical meaning can be linked with Slovenian counterpart *rastlina*, or German *Pflanze*, but has no connection to the meaning of a *production place* Slovenian *tovarna* or German *Fabrik*.

3 Description of the hierarchical dictionary as a graph

The most important property of the network of words is its organization as a connected cyclic hierarchical graph. The graph's elements are nodes (single words or clusters of words) and links between the nodes. The connected graph enables a continuous walk between any two nodes and is described as a sequence (path) of connected nodes. The graph is cyclic if it contains closed paths (cycles), i.e., paths that starts and ends on the same node) with all nodes on that path different (with exception of the closing node). Hierarchical graph has a single special node called top node N_{top} or root, distinguished from all others by defining the orientation of the graph and the walk directions within it. All paths between nodes must take one of the two directions: either *towards* the N_{top} (up) or *backwards* from it (down). Therefore, each node must have two list for connections (addresses), one to *up* and the other one to *down* connected neighbors, respectively. Similar to the N_{top} which is the last node for all *up* paths, at the end of any *down* path is always a node called *terminal node*, having no down addresses. In our case the terminal nodes are individual words.

The fact that the walk's path is not allowed to change direction assures the walk from any node will always reach either a *terminal* node (word) or *root* (N_{top}). Thus no walk keeping direction of moving could be captured in a cycle and thus end in an infinite loop. In the case of update or relocation of nodes checking the correct linkage of addresses of new connections prevents to generate infinite loops and self-referencing nodes. All the explained features of our graph offer the advantage of calculation the topological distance between the nodes. The topological distance D_{ij} between two nodes N_i and N_j has all four properties classifying it as a standard metric distance:

1) $D_{ij} > 0$ for all $i \neq j$
2) $D_{ij} = 0$ only for $i = j$
3) $D_{ij} = D_{ji}$, the distance is symmetrical, and
4) $D_{ij} \leq D_{ik} + D_{ki}$ triangle rule for any node k

To evaluate all topological distance D_{ij} between arbitrary two nodes N_i and N_j, one needs a complete connectivity matrix of order $(N_i \times N_j)$. For a graph containing approximately 10^5 nodes this means storing and handling the matrix of about 0.5×10^{10} distances. Fortunately, instead of keeping such large matrix only two connectivity *tables* one for keeping all *up* and the other one keeping all *down* connections from each node to its neighbors are needed. Using these two connectivity

tables it is straightforward to determine topological distance between any two nodes N_i and N_j or words i and j, respectively (Figure 1). The procedure is as follows:

1. generate a complete set $\{P_i\ (N_i,N_{top})\}$ of n_i paths from the leaf N_i to the N_{top},
2. generate a complete set $\{P_j\ (N_j,N_{top})\}$ of n_j paths from the leaf N_j to the N_{top},
3. pair-wise compare paths from both sets $\{P_i\ (N_i,N_{top})\}$ and $\{P_j\ (N_j,N_{top})\}$ to determine common nodes CN_k, k=1,… $n_i(n_j$-1)/2,
4. from the path set $\{P_k(N_i\text{-}CN_k\text{-}N_j)\}$ determine a set of paths lengths $\{\ l_k\ \}$, k=1,.. $n_i(n_j$-1)/2:
5. the shortest l_k is selected as the distance D_{ij} between nodes N_i and N_j:

$$D_{ij} = min\ \{\ l_k => P_k,(N_i\text{-}CN_k\text{-}N_j)\ \},\ \ k=1\ldots\ n_i(n_j\text{-}1)/2 \qquad /1/$$

4 Discussion

For highly flexible language like Slovenian, the parser is very important, because besides the grammatical analysis (tagging) it does the conversion of words into appropriate lexemes. Our system handles 149,000 Slovenian word-roots and about 3,100 different endings. The lemmas identified by the parser are handed to the search engine in the graph and finally the paths from lemmas to the top node N_{top} are reported. The output of the text analysis as given by our module is similar but not equal to the one obtained by another Slovenian parser by Amebis available on the ZRC portal ([11]). All tasks performed by the parser and by the search engine in the graph are executed *ab initio*, i.e. without any language corpus or internet connection.

The system we are describing can serve as a model how using a hierarchy of word clusters can be used for enhancing the information in any text. First, it provides grammatical information for and makes a lexical entry (lemma) for each word from the text and, second, it adds as many chains of relevant supplemental information about all nouns and verbs in the particular sentence. Each chain is a sequence of labels of the nodes (clusters of words) encountered during the *up* walk between the word and the N_{top}. Search algorithm finds *all* possible *up* walks from any encountered noun or verb to the N_{top}. (the reader can verify this part of the search engine on the link given in ([12])). Mostly, the labels are organized in self-explanatory manner using structure of keywords in which each keyword is itself a cluster label with the link to the particular cluster in the network. For example, the node labeled *property* (*human*) contains words each of which marks a *property of a human'* (*intelligence, beauty, greed, innocence,* etc.). On the other hand, the words in the cluster with the *same* two keywords, but ordered differently e.g., *human* (*properties*) a words assigning human being having the particular property (*miser, genius, liar,* etc.) are stored. Additionally, both words *human* and *property* are labels of other clusters. The cluster *property*, for example, contains 5,964 nouns with 14 sub-clusters named *property* ($keyword_i$), i = 1,…14. Each keyword of these clusters: property (*animal*), *property (human), property* (*number*), ... *property* (*object*) contains again cluster descriptors with keywords. Take for example the sub cluster *property* (*object*): *property* (*object* (*colour*)), *property* (*object* (*form*)), *property* (*object* (*price*)). At the end each $keyword_j$ represents a cluster with a smaller set of words.

Due to many ambiguities of Slovenian language no parser is perfect and therefore the errors are unavoidable. In the case of ambiguity our parser does not select one answer, but leaves both (or more) possible solutions on the output. For example, a very simple grammatically correct sentence in Slovenian language: *To je dobro za vas*, has two different interpretations, the first one, *That's good for you*, and the second one, *That's good for the village*. The appropriate meaning of this sentence can't be determined, until further explanation in one of the following sentences is given or the meaning is guessed from the previously learned context. Another similar example is a title of a well-known Slovenian short story entitled *Martin Krpan*, which introduces the name of the main character of the story. Surprisingly, the personal name Martin is again undistinguishable from the adjective Martin with the meaning *of belonging to the female named Marta*. As said before, such ambiguities are not rare in Slovenian language and therefore during the sentence analysis we prefer to leave all possibilities open for further analysis. In this way the user is warned about the possible mistakes and ambiguities, while the developers will recognise the cases that have to be dealt with, and handled for further improvements of the system.

5 The example

In order to show the procedure for the evaluation of the distance between two words /1/ an analysis of the sentence *Because the bridge on the violin has been broken, the strings have split up* (slov. *Ker je bila kobilica počena, so se strune strgale*) (Table 2). In the example the word *kobilica* having four very different meanings in Slovenian language (*locust, keel, diminutive of mare, and the bridge on the violin*) has been chosen deliberately to show the differentiation power of the topological distance evaluation between the nodes in the graph.

First, the tagging is made to provide lemmas for the input to the network. Second, the a list of all possible paths between the lemmas found and the N_{top} is

generated, and finally, according to the procedure /1/ the distance matrix between the nouns is evaluated from the obtained paths. The result is shown Table 2. Three of the common nodes CN_k on the paths are printed bold.

Table 3 shows the topological distance matrix between six words. In Slovenian language the word *kobilica* has four meanings. *the locust*, *the keel*, diminutive of *mare*, and *the bridge on the violin*. It is important to keep in mind that the word *kobilica*, although having four meanings is represented within the network with a single node, however, from this particular node four different paths are leading towards the N_{top}. All ten distances given in the distance matrix reflect the relations between the meanings of the words concerned reasonably good. The shortest distance of 3 nodes is between the word *kobilica* as *the bridge on the violin* and the *string*, respectively. Both words are in the same cluster labeled *part of (musical instrument)*, hence, the path: *kobilica(bridge) --> part of (musical instrument) --> struna (string)* has only tree nodes. The distance of 5 nodes makes the relation between three man-made objects complete. The distances *D*(*keel, bridge*) and *D*(*keel, string*) of 5 nodes is the same for both pairs reflecting the fact that all three involved objects are members of a relatively narrow group of man-made objects compared with others two words representing animals, locust and mare, respectively, which are part of the group *nature* (*animal kingdom*). The distance of 12 between the *locust* and the diminutive of *mare* is again consistent with distances between man-made objects because it can be interpreted that, although quite different, both 'animals' are in the group labeled *nature* (*animal kingdom*), but still quite far away. The largest distance of 18 nodes is associated with words between which the shortest path leads *via* the node *Noun.*

Ker; ker (*because*); cj.;
je; je (*is*); aux v. 3. prs. f/s past
bila; biti (*to be*); v. impf. n.refl. 3. prs. f/s past
kobilica; kobilica (*locus,keel,diminutive of mare; bridge on the violin*); f/s/1;
počena; počen (*broken*); adj. f/s/1;
, ločilo (separator); vejica (comma)
so; biti (*to be*); aux v. 3. prs. f/p past;
se; se (*oneself*); pron. mfn/sdp/124
strune; struna (*string*); f/p/14
strgale; strgati (*to split*); v. perf. refl. 3. prs f/p past;
. ločilo (separator); pika (full-stop)

a) kobilica: insect; insect (*pterygota*)*; insect* (*arthropoda*)*; non-vertebra;* ***nature*** **(*animal kingdom*)***; nature* (*live*)*; Nature; Noun;* N_{top}. (length = 10).

b) kobilica: mare; horse (*animal*(*general*))*; horse* (*animal*)*; animal* (*domestic*)*; animal* (*property*)*;* ***nature*** **(*animal kingdom*)***; nature* (*live*)*;Nature; Noun,* N_{top}. (length = 11).

c) kobilica: part of (*musical instrument*)*;* ***part of*** **(*specific device*)***; product* (*part of*)*; product* (*one object*)*; product* (*material*)*;* ***product*** **(*man made*)***; Product; Noun;* N_{top}. (length = 10).

d) kobilica: part of (*vessel*)*;* ***part of*** **(*specific device*)***; product* (*part of*)*; product* (*one object*)*; product* (*material*)*;* ***product*** **(*man made*)***; Product; Noun;* N_{top}. (length = 10).

a) biti: to exist (*person*)*; to be* (*to exist* (*specific way*))*: To be; Verb;* N_{top} (length = 6).

a) struna (string): part of (*musical instrument*)*;* ***part of*** **(*specific device*)***; product* (*part of something*)*; product* (*one object*)*; product* (*material*)*;* ***product*** **(*man made*)***; Product; Noun;* N_{top} (length = 10).

a) strgati (*to grate*)*: to remove* (*surface*)*; to remove; to put somewhere; verbs of doing; Verb;* N_{top}. (length = 7).

b) strgati (*to split up*)*: to divide into parts; to divide; to work with parts; verbs of doing; Verb;* N_{top}. (length = 7).

Table 2 Output of the analysis of the sentence *Because the bridge on the violin has been broken, the strings have split up (translated to English*). Below the tagged words the chains of nodes from the graph are listed as they appear in the analysis. Some examples for the abbreviations are given in parentheses: e.g. for verbs, (v impf, n.refl. 3 prs. f/s past) means: verb imperfect, not reflexive, 3rd person, female, singular, past tense; for nouns and adjective, (f/s/1) and (adj f/s/1), respectively, it mans: female, singular, nominative, etc.

	Wordi or node N_i	*kobilica* (*locust*)	*kobilica* (*keel*)	*kobilica* (dimin. *mare*)	*kobilica* (violin *bridge)*	*struna* (*string*)
1	*kobilica* (engl. *locust*)	0	18	12	18	18
2	*kobilica* (engl. *keel*)		0	18	5	5
3	*kobilica* (engl. dim. of *mare*)			0	18	18
4	*kobilica* (engl. violin *bridge)*				0	3
5	*struna* (engl. *string*)					0

Table 3. Distance matrix ||D|| of topological distnces between five words. Matrix elements $D_{ij} = D(N_i N_j)$. Using the given paths in the Table 2 and procedure /1/, the reader can verify the values of topological distances in the matrix.

6 Conclusion.

Neither the presented network, nor the presented model for extracting broader information from the text, is the final products. The discussed example and hierarchical network of words are both simple and small parts of the possibilities that can be accomplished by the use of an exhaustive and therefore much more complex network (graph) of word-meanings. Still a lot of improvements are waiting to be implemented.

Although in the presented network more than 60,000 lexemes are hierarchically linked with more than 5,000 clusters emphasizing a wide variety of properties and features of words and meanings, it is not the number of words that is a limiting factor, but rather the absolute number of nodes (clusters of words with different features) and the number of nodes and topological paths to which each word is linked. This is the issue that should be of first concern and improvement. One should add not only more clusters presenting still larger variety and number of properties, features, and/or meanings, but as well clusters of features that the words *do not* have or represent, or clusters of words with exclusive-or properties, etc. The compilation of such network will require much more time than it has been spent for the building the present one, although it takes approximately eight man-years to reach the present size of the existing one.

Some critics are afraid that such knowledge bases has arbitrary structure, because the choice of clustering and selection of features and properties are subjective and no objective criteria exist how to select and link clusters of words according to their meanings and features. The argument that such a hierarchy will always be subjective is true, indeed, but so is the human mind.

An additional challenge is the implementation of the English vocabulary into the basic skeleton of the existing hierarchical skeleton of Slovenian words and their meanings. This task is almost impossible to accomplish successfully without permanent assistance of a native speaker. Therefore, the skeleton of the described hierarchy is free to any one interested in it and available from the author.

The topological distance measure in the graph is an additional challenge. The present procedure for distance calculation /1/ which does not take into consideration the distance between the common nodes (CN) and the N_{top} what sometimes distorts the relations within the distances among several words or word clusters. Therefore, a kind of weighting procedure that would considers the absolute position of the CN were the paths meet to the N_{top} should be introduced.

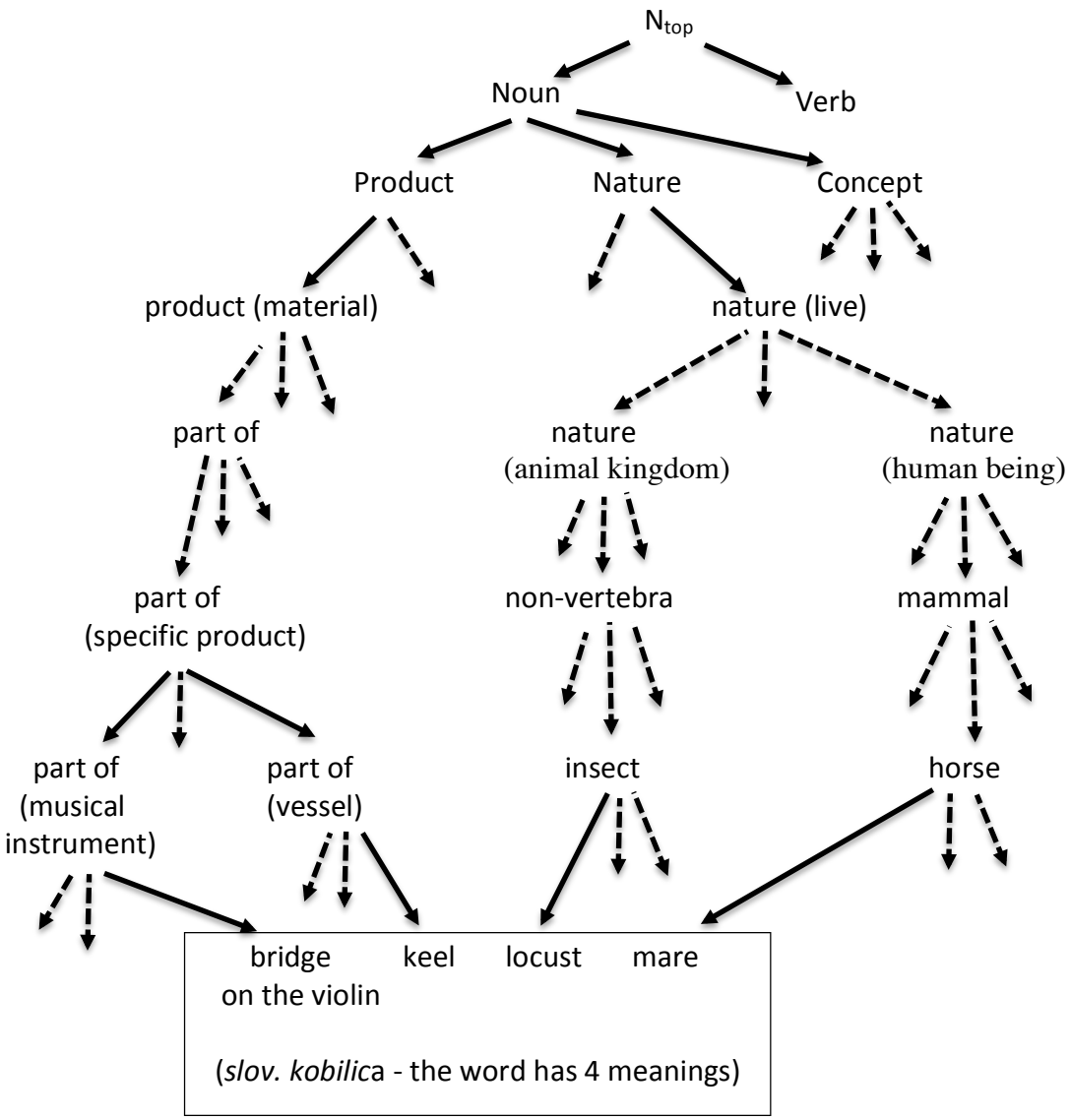
The time needed for the complete analysis (tagging of all words and finding all shortest paths from the words to the N_{top} in the text of 500 sentences with about 6,000 words requires less than 2 second. However, with the increasing vocabulary and complexity (number of links) within the lexical graph the speed of the search engine could be a problem. In the present status of the size the vocabulary and the complexity of the links in the collection the speed of the analysis is acceptable.

Acknowledgement

The author wish to thank National Institute of Chemistry, Ljubljana, for providing him with the facilities to work at the Institute as the Research Emeritus.

References

1. N. Trinajstič, *Chemical Graph Theory*, CRC Press, Boca Raton, 2. Edition, 2000,
2. J. Zupan, A. Lajovic, *Pomenska mreža samostalnikov in glagolov*; v Obdobja 34, Slovnica in slovar – aktualni jezikovni opis, Ed., M. Smolej, Filozofska Fakulteta, Ljubljana, 2015, pp. 871-878.
3. B. Levin, English Verb Classes and Alternations, ,
4. F. Dorenseiff, *der deutsche Wortschatz nach Sach-gruppen*, 8. Edition, Ed. U. Quasthoff, W. de Gruyter, Berlin, 2004.
5. Visuword™, *On-line graphical dictionary and thesaures*, https://visuwords
6. D. Fišer, J. Novak, *Visualizing sloWNet*, v Proc. Conf. on Electronic Lexicography in the 21st Century: New applications for new users (eLEX2011). Bled, Slovenia, 9-12 November 2011
7. J. Zupan, *Hierarhična mreža slovenskih glagolov*, v Obdobja 30, Interdisciplinarity in Slovene Studies, Filozofska Fakulteta, Ljubljana 2011, pp. 551-557.
8. J. Zupan, *Koncept mrežnega pomenskega slovarja slovenskih besed*, Jezik in slovstvo, 54, (3-4), 2009, pp. 139-151
9. J. Zupan, A. Lajovic, PMSG – Network of Slovenian verbs, web address: http://pmsg.zrc-sazu.si.
10. J. Zupan, *Pomenska mreža slovenskih glagolov*, Založba ZRC SAZU, 2013, pp. 31-51
11. *Oblikoslovni označevalnik za slovenski jezik*, Amebis, d.o.o. Kamnik, Inštitut Jožef Stefan, Univerza v Ljubljani, ZRC SAZU, Trojina, Zavod za uporabno slovenistiko, 2008-2013, konzorcij projekta Sporazumevanje v slovenskem jeziku: link to the network: http//www. oznacevalnik.slovenscina.eu
12. J. Zupan, A. Lajovic; *PMSB, Pomenska mreža slovesnkih besed*, link to the network of meanings of Slovenian words: http://mreza.andrej.ad-vega.si.



.

Figure 1. A simplified part of the discussed network of words showing essential features of a cyclic bi-directional graph. Each label represents a node. A cycle is a path that starts and ends on the same node. From the word kobilica having 4 meanings in Slovenian language, six cycles can be drawn to calculate six distances between all four meanings. Because the graph is 2-directional only the paths in 'up' or 'down' to the N_{top} (opposite to arrows) or to the terminal nodes (words, along arrows)), respectively, are allowed. The cycles are detected *via* the common nodes CN_k on the paths

.