# HYDRA FOR WEB: WORDNET ONLINE EDITOR

**Borislav Rizov, Tsvetana Dimitrova**
*Institute for Bulgarian Language – Bulgarian Academy of Sciences*

**Abstract.** The paper presents the CRUD functions of the Hydra for Web system for work on lexical-semantic databases with relational structure similar to the structure of WordNet. It supports functionalities for editing of relational data, simultaneous access of multiple users, parallel data visualisation. Hydra for Web has been used for the development of the Bulgarian wordnet.

*Keywords:* content management system; language data; lexical-semantic networks; Wordnet

## 1. Introduction

WordNet[1] is a lexical-semantic database firstly created for English – the Princeton WordNet (Miller, 1995; Fellbaum, 1998) where synonymy is the main relation between words, and synonyms (termed 'literals') are organised in unordered sets called synonym sets (synsets) that are interlinked via conceptual-semantic and lexical relations.

Wordnet encodes data into relational format and needs flexible tools to give the users easy access to data editing and visualisation. One such tool is Hydra for Web that is at focus of the present paper. It is a system application for working with complex relational data of wordnet (including parallel data between two or more wordnets). It integrates an editing functionality and a simple interface that keeps the structure of a synset with all the relations integrated into one hierarchical structure.

The paper contains a discussion on: the wordnet and its structure (section 2); the Hydra for Web system's interface and functionalities (section 3); and a brief overview of its applications (section 4).

## 2. WordNet

WordNet is a large lexical database of English (started as the Princeton Wordnet, cf. Miller, 1995) which contains nouns, verbs, adjectives and adverbs that are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept (synsets are considered to reflect (psycho)linguistic concepts, as claimed by Miller (1990). In the last decades, wordnets for over 43 languages have been

developed and among them, the Bulgarian wordnet (BulNet) (Koeva et al., 2004; Koeva, 2010) which started within the project BalkaNet – a Multilingual Semantic Network of the Balkan Languages (Stamou et al., 2002) covering five Balkan languages – Bulgarian, Greek, Romanian, Serbian, Turkish, plus Czech. In the recent years, the Bulgarian WordNet has been expanded to over 120,000 synsets (of which over 80,000 are manually validated). While the Princeton WordNet (PWN) includes only open class words, i.e., nouns, verbs, adjectives, adverbs, synsets in the Bulgarian wordnet are distributed into nine parts-of-speech – nouns, verbs, adjectives, adverbs, pronouns, prepositions, conjunctions, particles, and interjections (function words were added within the work on the development of the Bulgarian Sense-Annotated Corpus where every word is linked to a corresponding synset (Koeva et al., 2011).

Synonymy is the main relation between the words, and synonyms ('literals') are organised in unordered sets (synonym sets or synsets). A synset obligatorily contains literals (at least one, in the examples below, these are the sets of words into curly brackets), a definition, an identification number (**ili** – interlingual lexical index (Vossen, 2002) and information about the part-of-speech (**pos**). It often contains usage examples (one or more phrases or short sentences) to illustrate the use of the concept.

Synsets are interlinked via conceptual relations – hypernymy/hyponymy, antonymy, meronymy, holonymy, etc., morphosemantic relations (agent, event, result, material, location, etc.) (for detail, see Miller, 1995; Fellbaum, 1998; Fellbaum et al., 2009}. Nouns and verbs are obligatorily linked to hypernyms, as in Ex. 1[2] where *{летище:2; летателно поле:1}* // *{airfield:1; landing field:1; flying field:1; field:6}* is a hypernym (relation **hypernym**) of the synset *{аерогара:1; аеропорт:1; летище:1; летищен комплекс:1}* // *{airport:1; airdrome:1; aerodrome:1; drome:1}* where each member (a single word or a phrase/multiword expression) is a literal. Synsets are also interlinked via conceptual relations such as meronymy/holonymy (relations **more_part, holo_part**) as with *{хангар:1}* // *{airdock:1; hangar:1; repair shed:1}* in Ex. 1, antonymy (relation **antonym**) (for adjectives and adverbs), among other.

**Ex. 1:**

**Synset:** bg – {n: аерогара:1; аеропорт:1; летище:1; летищен комплекс:1} // { n: airport:1; airdrome:1; aerodrome:1; drome:1}

**definition:** *летателна зона заедно с комплекса от наземни съоръжения (писти, хангари, сгради) за излитане, кацане, престой и обслужване на самолетите и осигуряване на редовния превоз на пътници и товари по различни въздушни линии*

**pos:** n **ili**: eng-30-02692232-n **semantic class:** noun.artifact

**hypernym:** bg – {n: летище:2; летателно поле:1} // {n: airfield:1; landing field:1; flying field:1; field:6}

**hyponym:** bg – {n: хелипорт:1} // {n: heliport:1}
**mero_part:** bg – {n: хангар:1} // {n: airdock:1; hangar:1; repair shed:1}
**mero_part:** bg – {n: летищен терминал:1} // {n: air terminal:1; airport terminal:1}
**mero_part:** bg – { n: контролна кула:1} // {n: control tower:1}

Adjectives are linked to other synsets via derivational relations and relations such as **similar_to** and **also_see** (to their near synonyms), as in Ex. 2.

**Ex. 2:**
**Synset:** bg – {a: далекоглед:}1 // {a: farsighted:2; presbyopic:1}
**definition:** *който вижда ясно далечни обекти, но понякога вижда мъгляво близки обекти (при хората обикновено след 45-годишна възраст*
**pos:** a **ili:** eng-30-02157399-a **semantic class:** adj.body
**also_see:** bg – {a: предвидлив:1} // {a: provident:1}
**antonym:** bg – {a: късоглед:1; миопичен:1} // {a: nearsighted:1; shortsighted:3; myopic:2}
**eng_derivative:** bg – {n: хиперопия:1; хиперметропия:1;далекогледство:1} // {n: hyperopia:1; hypermetropia:1; hypermetropy:1; farsightedness:2; longsightedness:1}
**eng_derivative:** bg – {n: пресбиопия:1; старческо далекогледство:1} // {n: presbyopia:1; farsightedness:3}
**similar_to:** bg – {a: хиперметропичен:1; далекоглед:3} // {a: hyperopic:1; hypermetropic:1}

Adverbs are often linked via **derived** / **derivative** relation to the adjective from which they are derived, as exemplified in Ex. 3. Adverbs in BulNet have an obligatory relation **category_member** via which they are linked to synsets that define specific concepts of time, place, manner, frequency, etc. (e.g., *{красиво:1}* is linked via **category_domain** to *{начин:3; стил:2; маниер:1}* in Ex. 3).

**Ex. 3:**
**Synset:** bg – {b: красиво:1} // {b: beautifully:1; attractively:1}
**definition:** *характеризира посоченото в изказването като будещо приятни естетически усещания*
**pos:** b **ili:** eng-30-00242006-b **semantic class:** adv.all
**category_domain:** bg – {n: начин:3; стил:2; маниер:1}
**antonym:** bg – {b: неатрактивно:1; непривлекателно:2} // {b: unattractively:1}
**derived:** bg – {красив: 1; прекрасен:2} // {beautiful:1}
d**erived:** bg – {привлекателен:1} / {attractive:1}

Synsets can also be connected via morphosemantic relations (**agent, event, result**, etc., see Miller, 1995; Fellbaum, 1998; Fellbaum et al., 2009), as in Ex. 4 where *{пушене:1; тютюнопушене:1} // {smoke:2; smoking:1}* is event of

*{пуша:2; изпушвам:1; изпуша:1} / {smoke:3}*. In Ex. 4, relations on the level of literal are also exemplified with the literal *{пушене:1}* being derivationally related to the literal *{пуша:2}* of the corresponding verb synset via the literal derivational relation **without_suffix** – it basically means that there is no derivational suffix in the verb form *пуша* (thus, it is marked as being 'without suffix') (the approach to introducing a range of derivational relations on literals are described in (Dimitrova, et al., 2014).

**Ex. 4:**
**Synset:** bg – {n: пушене:1; тютюнопушене:1}
**definition:** *актът на пушене, поемане през устата на дим от тютюн или други субстанции*
**literal:** {пушене:1}
**without_suffix:** {пуша:2}
**literal:** {тютюнопушене:1}
**usage:** Пушенето е позволено само навън.
**usage:** Пушенето усмърдява стаята.
**pos:** n **ili:** eng-30-00834636-n **semantic class:** noun.act
**hypernym:** bg – {n: дишане:1}
**hyponym:** bg – {n: пуфтене:2}
**bg_derivative:** bg – {v: пуша:2; изпушвам:1; изпуша:1}
**is_event_of:** bg – {v: пуша:2; изпушвам:1; изпуша:1}
**mero_part:** bg – {n: дръпване:2; всмукване:1; дърпане:3; пухтене:2}

Additionally, each synset is classified by a semantic primitive (Miller 1993 et al., Fellbaum et al., 2009}. Nouns are organised into 25 semantic classes (noun.person, noun.animal, noun.plant, noun.event, noun.act, noun.body, noun.artifact, etc.) while verbs are classified under 15 primes (verb.stative, verb.communication, verb. change, verb.cognition, verb.cognition, verb.body, etc.). In the Princeton WordNet, adjectives are classified into two larger classes: descriptive adjectives and relational adjectives; plus an additional class of adjectival participles (Fellbaum, 1993), while other wordnets have introduced more detailed classifications (including the Bulgarian wordnet, cf. Stefanova 2016; Stefanova & Dimitrova, 2017).

## 3. Hydra for Web: User Interface and Functionalities

Hydra for Web is based on the notion of wordnet as a relational structure organised around a set of objects that are interlinked via a set of binary relations. The objects are of three types – Synset, Literal and Note. The Literals (i.e., the words) in a synset are connected with it via a relation called **literal**. The Notes objects represent the textual data in wordnet – usage examples and notes. In a way similar to literals, every usage example is connected to its synset via the relation **usage**.

Hydra for Web started as a tool intended to show parallel wordnet data where wordnets with the same identification can be visualised in parallel (for this purpose, wordnets use the ILI – interlingual lexical index (Vossen, 2002). The editor was further developed for online editing of the wordnet data. Hydra for Web (available at http://dcl.bas.bg/bulnetedit/ - literals (and synsets) that have not been validated yet and are not part of the validated BulNet database are dimmed) is a single page web application that uses as backend the API of the open source modal logic tool for wordnet development Hydra (Rizov, 2008; downloadable at: http://dcl.bas.bg/hydra/). It is built with Node.js (Node.js® is a JavaScript runtime: https://nodejs.org/.) and Express (Web application framework for Node.js, http://expressjs.com/). The wordnet data retrieval is made by means of the Wordnet Service. It is mobile-friendly on a small width (mobile), where the panels are ordered successively.

The navigation bar (as seen on: http://dcl.bas.bg/bulnetedit/) has a dropdown menu for switching between the wordnets the user wants to work on. It contains modes such as BulNet vs. PWN, BulNet vs. RoWN and RoWN vs. PWN by default but through a modal dialog the user can enable / disable additional wordnets. The user can search for a word in Bulgarian and see the corresponding synsets in English and Italian, for example. The tool allows users to work on any wordnet that is in the database. The interface is currently available in English, Bulgarian, and Romanian but it is possible for other languages to be added.

### 3.1. Search

The search system provides results in all of the available languages (restricted by the selection of the user) – in addition to the Princeton WordNet (PWN) 3.0, and the Bulgarian WordNet (BulNet), the database currently contains over 20 other wordnets. The tool allows for searching into databases of different language wordnets with a single query. The selected result by the user is propagated to the right hand side visualiser(s). Hydra for Web supports two visualisation modes:

– Single mode – one visualiser where what you select is what you see;

– Bilingual mode – two visualisers – where you see the correspondences of the selected synset in the mode's languages.

Every object visualisation is recursive in a sense that every relation (**hypernym**, **holo_part**, etc.) that leads to other object (i.e., synset) is expandable in the same way as the root one. The data in objects like pos, ILI, etc., are available immediately, while the relations are loaded by means of AJAX query, but asynchronously without blocking the UI.

The tool allows searching for an exact match of a word string – a single word such as *{чай}*, or a multiword unit, e.g., *{кутия за чай}*, or a non-exact match search which returns any synset where the searched word is found (e.g., a search for *{чай}* returns 22 synsets including *{черен чай:1}*, the adjective *{с чаен аромат:1, с аромат на чай:1}*, etc.).

Although the three types of objects are fully-fledged, the search panel returns all the synsets that contain a literal matching the search query.

The search input is enhanced with autocomplete (with prefix match) as shown on Fig. 1 for a search for the word *кафе* 'coffee' (up to to 10 elements are shown at once in a list).

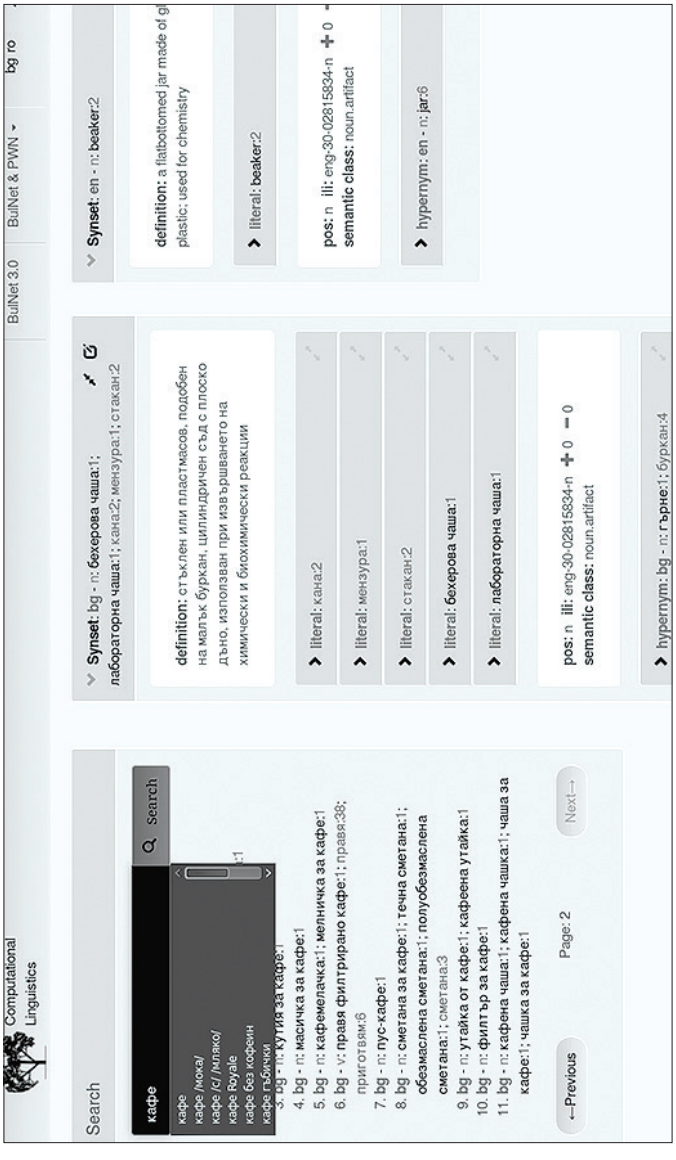


**Fig. 1: Hydra for Web Editor – Search**

The search returns a paginated list with the respective synsets in the database, as shown on Fig. 1 for the non-exact match search for *кафе* 'coffee' which returns 41 different synsets from the Princeton WordNet database (the exact match search returns only 7 synsets). The results shown at once are limited to 30 synsets in a list below the Search input but the user can access all the synsets found by using the button Next and Previous to browse between the results (pages).

The tool shows the status of the data (literals) – literals that have been validated by an expert, are visualised in the standard color, while those that have not been validated yet, are dimmed (muted).

To limit the results shown, the search respects word (string) boundaries, i.e., the user can search only for whole words but not parts of the words.

### 3.2. Editing a synset

The online editor Hydra for Web allows to:

– Edit object's data. Some of the fields require free text like definition, while others are with predefined value list – the part-of-speech.

– Add object (literals and notes are added by button clicks in the parent objects).

– Delete object.

– Add binary relation (i.e., hypernym) between existing objects.

A synset can be edited by clicking on the top right-edge Edit button of the panel to put the linguistic unit (Synset, Literal or Note) panel in Edit mode – the data visualisation controls are replaced with those for editing.

The Edit panel for a synset consists of subpanels for elements which are part of the synset – at least four, as shown on Fig. 2: the set of literals constituting a synset; the definition; the literals visualised as a list – each literal can be edited as an independent object with its own structure; and information that is unique to the current synset only – part-of-speech (pos), ILI, sentiment values according to SentiWordNet (Esuli & Sebastiani, 2006), semantic class. Other elements that can be visualised are: **usage**, **snote**, relations such as hypernym, hyponyms, derivational relations, morphosemantic relations, and others. Fig. 2 shows a synset that has not been edited yet. The information in Bulgarian as seen here, is not accessible publicly but only to the editors with the appropriate rights. The information has been automatically translated and added into the database to help the editors.

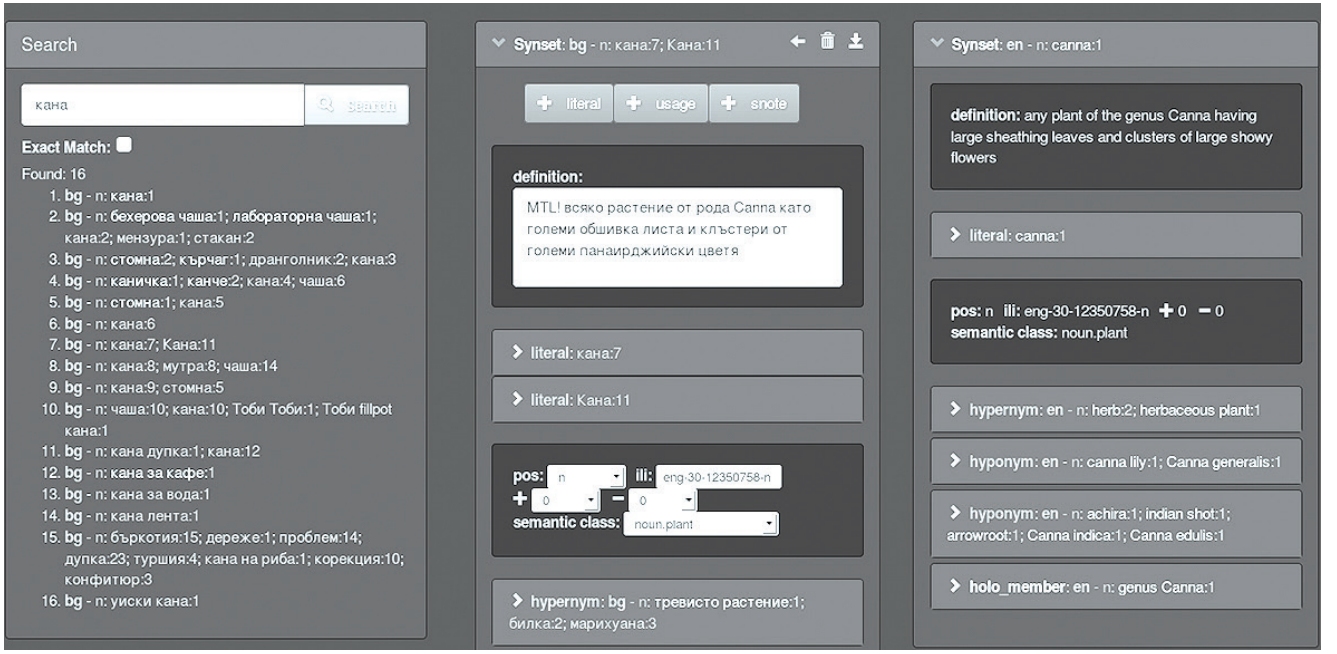


**Fig. 2: Hydra for Web Editor – Edit Panel**

From top to bottom, the following elements are part of the editor panel (for Synset object in the example) as shown on Fig. 2:

1. Panel header – textual representation of the synset – all the literals to the left, followed by buttons for canceling (the green arrow sign), deleting (the 'bin' sign), and saving the synset.
2. Three buttons for adding (with the plus sign) **literal**, **usage** and **snote** relations (and new objects) of the synset.
3. The definition.
4. The literals ordered in a list. Each literal can be edited independently by clicking on the Edit button and opening an Edit panel which is much like the Editor panel of the parent synset. By clicking on the literal – without opening the Edit panel – the user can view the whole information about the literal at hand (word, lemma, status, and **lnote** if available plus the entire synset – with all the literals – it pertains to).
5. Information about: pos, ILI, sentiment values according to SentiWordNet, and semantic class. All values of these categories are editable – pos, SentiWordNet values, and semantic class are available as a list with fixed values.

The synsets to which a currently edited synset is linked to via a relation (hypernym, hyponym, etc.) are given as a list after the subpanel (5) and each of the linked synsets can be edited further on its own.

### 3.3. Linking

The Linguistic Units (LU) can be connected by introducing a relation between the two connected LUs. It is accomplished by means of a Wizard. To start, the user clicks on the Connect button to the left of the Edit button on the unit panel. The procedure requires the following steps:

Step 1: A new Select Relation panel is opened to replace the Search panel. The new panel offers a list of all the relations available for the selected type of LU – as seen on Fig. 3.


**Fig. 3: Hydra for Web Editor – Select Relation**

Step 2: The target LU of the relation is shown via a Search panel identical to the main Search panel. The search returns a list of synsets to be linked to the selected synset. Fig. 4 shows a selection of the **is_agent_of** relation that has to link the synset *{ковач:1; железар:1}* 'blacksmith' to the synset *{кова:2}* 'forge, hammer'. The selection of a target synset from the searched for list in the Search panel shows the whole synset below the list in the Search panel. If this is the intended synset, the user clicks on the button Connect and the link is visualised on the panel to right.

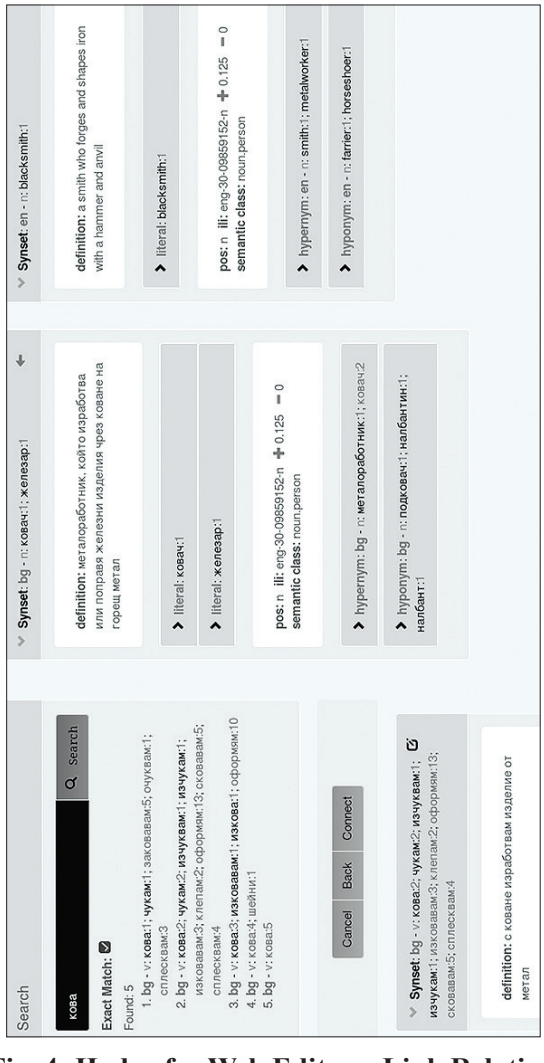


**Fig. 4: Hydra for Web Editor – Link Relation**

### 3.4. Concurrent Editing

All modified data is propagated to the other connected users immediately by means of notifications by the wordnet server. In case of a conflict (the same object is edited by more than one user), the last user is responsible for merging the data. When receiving a notification that some data is in edit mode, Hydra puts it in merge mode.

Hydra for Web is freely accessible to all. Anonymous users have access to view BulNet, PWN, RoWN, SlovakWN and ItaWN for the time being. Additionally, the system is enhanced with user management with the following privilege options for every given language/wordnet:

– The wordnet is unavailable to the user.
– View: The user can search and browse this wordnet.
– Edit: The user can edit the data and relations in this wordnet.

## 4. Applications

Hydra for Web can be used for queries and viewing and for development of wordnets or any lexical resource that has relational structure similar to the structure of the wordnet. Parallel data can be used for comparative lexical and other linguistic studies including translation studies (as the synsets are also edited with a view to being interpreted as translation equivalents, cf. Angelov & Lobanov, 2016), for improving machine translation systems (Kim et al., 2002; Salam, 2009).

It can be used also as a multilingual dictionary especially since the list of results (single words and multiword units) returned also contains information about other (synonym) words and the part-of-speech of the resulting words.

It can be used also in language teaching including in foreign language teaching (possible applications in this area are discussed in (Leseva et al., 2016).

**NOTES**

1. 'WordNet' is used for the Princeton WordNet (WordNet for English); other lexical-semantic networks that have being developed following the model of the Princeton Wordnet are termed 'wordnet(s)' (e.g., the Bulgarian wordnet).
2. The examples in the article are extracted from the Bulgarian wordnet - the parallel English synsets follow the Princeton WordNet (as can be viewed at: http://dcl.bas.bg/bulnet/). The numbers of the literals (words) are arbitrarily applied.

**REFERENCES**

Dimitrova, T., Tarpomanova, E. & Rizov, B. (2014). Coping with derivation in the Bulgarian Wordnet (pp. 109 – 117). In: Orav, H., Fellbaum, C. & Vossen, P. (Eds.). *Proceedings of the Seventh Global Wordnet Conference (GWC'2014*), Tartu, Estonia. Stroudsburg (PA): Association for Computational Linguistics.

Fellbaum, C. D. (ed.) (1998). *Wordnet: An electronic lexical database*. Cambridge, MA: MIT Press.

Koeva, S. (2010). Bulgarian wordnet – current state, applications and prospects (pp. 120 – 132). In: *Bulgarian-American Dialogues.* Sofia: Academic Publishing House.

Koeva, S., Tinchev, T. & Mihov, S. (2004). Bulgarian wordnet – structure and validation. *Romanian Journal of Information Science and Technology*, 7(1 – 2), 61 – 78.

Miller, G. A. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography* 3, 4 (Winter 1990), 235 – 312.

Miller, G. A. (1995*). Wordnet: A lexical database for English. Communications of the ACM,* November 1995, 38(11), 39 – 41.

Rizov, B. (2008). Hydra: A modal logic tool for wordnet development, validation and exploration (pp. 1523 – 1528). In: *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08*), Marrakech, Morocco.

Rizov, B. (2014). Hydra: A Software System for Wordnet (pp. 142 – 147) In: Orav, H., Fellbaum, C. & Vossen, P. (Eds.). *Proceedings of the Seventh Global Wordnet Conference (GWC'2014),* Tartu, Estonia. Stroudsburg (PA): Association for Computational Linguistics.

Vossen, P. (2002). WordNet, EuroWordNet and Global WordNet. *Revue française de linguistique appliquée,* 1/2002 (Vol. VII), 27 – 38.

Leseva, S., Stoyanova, I., Todorova, M. & Koeva, S. (2016). Language technologies and resources – new advances in Bulgarian language teaching (the Bulgarian Lexical Semantic Network BulNet and the Bulgarian National Corpus). *Bulgarian Language and Literature,* 4 (2016), 377 – 393. [Лесева, С., Стоянова, И., Тодорова, М. & Коева, С. (2016). Езиковите технологии и ресурси – нови перспективи в обучението по български език (Българската лексикално-семантична мрежа БулНет и Българският национален корпус). *Български език и литература*, 4 (2016), 377 – 393.]

Stamou, S., Oflazer, K., Pala, K., Christoudoulakis, D., Cristea, D., Tufis, D., Koeva, S., Totkov. G., Dutoit, D. & Grigoriadou, M. (2002). BalkaNet: A multilingual semantic network for the Balkan languages (pp. 21 – 25). In: *Proceedings of the International Wordnet* Conference, Mysore, India.

Fellbaum, C., Osherson, A. & Clark, P.E. (2009). Putting semantics into WordNet's "morphosemantic" links (pp. 350 – 358). In: *Proceedings of the Third Language and Technology Conference*, Poznan, Poland. [Reprinted in: Responding to information society challenges: New advances in human language technologies. Springer Lecture Notes in Informatics, vol. 5603.]

Miller G., Beckwith R., Fellbaum C., Gross D. & Miller K. (1993). Introduction to WordNet: An on-line lexical database. *Five papers on WordNet. Princeton University* (1993) http://wordnetcode.princeton.edu/5papers.pdf [19/06/2017]

Stefanova, V. 2016. Classificational model of adjectives in the Bulgarian wordnet (pp. 51 – 57). In: *World is Word, Word is World*. A volume of papers from the Jubilee National Conference with International Participation dedicated to 25th anniversary of the Philological Faculty of the South-West University "Neofit Rilski" (06.10. – 07.10.2016). Blagoevgrad: University Press "Neofit Rilski". [Стефанова, В. (2016). Класификационен модел на прилагателните имена в Българския Уърднет (с. 51 – 57). В: *Светът е слово, словото е свят*. Сборник с доклади от юбилейна национална конференция с международно участие, посветена на 25 години на Филологическия факултет (06.10 – 07.10.2016 г.). Благоевград: УИ „Неофит Рилски".]

Stefanova, V. & Dimitrova, T. (2017). Classification of adjectives in BulNet: Notes on an effort. In: Piasecki, M. & Bond, F. (Eds.). *Proceedings of the Challenges for Wordnets Workshop within the First International Conference, LDK 2017*, Galway, Ireland, June 19 – 20, 2017. (in press)

Esuli, A. & Sebastiani, F. (2006). SentiWordNet: A publicly available lexical resource for opinion mining (pp. 417 – 422). In: *Proceedings of 5th Language Resources and Evaluation Conference (LREC'2006)*. European Language Resources Association.

Angelov, K. & Lobanov, G. (2016). Predicting translation equivalents in linked WordNets (pp. 26 – 32). In: *The 26th International Conference on Computational Linguistics (COLING 2016*). Osaka, Japan.

Salam, K. Md. (2009). *Independent study report: Improving example based English to Bengali machine translation using WordNet*. Diss. BRAC University, 2009.

Kim, Y., Chang, J. H. & Zhang, B. T. (2002). Target word selection using WordNet and data-driven models in machine translation (pp. 607 – 607).

In*: Pacific Rim International Conference on Artificial Intelligence*. Springer Berlin Heidelberg.

Koeva, S., Leseva, S., Rizov, B., Tarpomanova, E., Dimitrova, T., Kukova, H. & Todorova, M. (2011). Design and Development of the Bulgarian Sense-Annotated Corpus (pp. 143 – 150). In: *Las tecnologías de la información y las comunicaciones: Presente y futuro en el análisis de córpora. Actas del III Congreso Internacional de Lingüística de Corpus*. Valencia: Universitat Politècnica de València.

✉ [1] **Mr. Borislav Rizov,** [2] **Dr. Tsvetana Dimitrova, Assist. Prof.**
Intstitute for Bulgarian Language "Prof. Lyubomir Andreychin"
Bulgarian Academy of Sciences
52, Shipchenski Prohod Blvd., bl. 17
Sofia, Bulgaria
Email: [1] boby@dcl.bas.bg; [2] cvetana@dcl.bas.bg