

ПРЕГЛЕД И ПРОУЧВАНЕ НА СЪЩЕСТВУВАЩИ СИСТЕМИ ЗА ИЗКУСТВЕН ИНТЕЛЕКТ

Мария Жекова

Пловдивски университет „Паисий Хилендарски“

OVERVIEW AND EXPLORATION OF EXISTING ARTIFICIAL INTELLIGENCE SYSTEMS

Mariya Zhekova

Plovdiv University “Paisii Hilendarski”

One of the tasks of computational linguistics is the creation of software and applications to facilitate access to and use of large amounts of information. Natural language processing can create interfaces that allow users to interact more intuitively with data, and to access information by entering queries using their native language. This article provides an overview of the history of information systems with a natural language interface to a database. Existing systems of artificial intelligence present knowledge and reflect the conceptual basis of the organization of human memory as well as its flexibility and visibility. The publication is part of a study devoted to the application of the natural language interface to e-learning.

Keywords: natural language understanding, artificial intelligence, natural language processing, data/information processing

Дългосрочно желание от страна на компютърните потребители е да минимизират пропагата в комуникацията между компютъра и човека. Потребителските интерфейси на естествен език (ИЕЕ) са вид компютърно-човешки интерфейси, където езикови форми като глаголи, фрази и клаузи действат като параметри в образуването на заявките към база данни за извличане на информация. Според Рукшън, Прашанти, Синатамби (Rukshan, Prashanthi, Sinnathamby/-Рукшън, Прашанти, Синатамби 2013) извличането на информация от базата данни изисква задълбочени познания за технически езици като Structured Query Language. За разлика от това интерфейс на естествен език за работа с база данни би позволил да се задават въпроси, които да бъдат формулирани на родния език на потребителя. Следователно информационни системи с естественоезиков интер-

фейс биха били по-подходящи за случайни потребители, тъй като няма да е нужно потребителят да изучава език, чрез който да комуникира със системата.

ИЕЕ трябва да е в състояние да разбере човешкия език след въвеждане и да го преведе на езика, който базите данни разбират, за да се получи информация. Адекватното представяне на вътрешната структура е ключът към постигането на добри решения на реални проблеми.

Машинното обучение е мощна форма на изкуствен интелект. Популярни потребителски приложения като Alexa, Siri, Netflix, Google Maps, Facebook новините и автопилота на Tesla са много успешни в овладяването на силата на машинното обучение да разпознават модели в огромни количества от данни (подобно на други нововъзникващи приложения за откриване и реагиране на заплахи за киберсигурност, автоматизация на процесите, управление на мрежова инфраструктура, застрахователно ценообразуване и търговия с акции, облигации и други финансови инструменти).

Но човешкият интелект е много повече от просто разпознаване на модели, а изкуственият интелект е много повече от машинно обучение. Ефективността на приложенията за машинно обучение е ограничена до наличието на правилен набор от пълни, коректни, ясни и точни данни, като се разчита на човешки специалисти по бази от знания да конструират, настройат и усъвършенстват моделите и невъзможността на тези еднократно създадени модели да се обобщят за всяка предметна област, за всички бизнес процеси, функции и дори географски региони.

Видове системи за изкуствен интелект

Една от задачите на компютърната лингвистика е създаването на софтуерни програми и приложения, които да улеснят достъпа до огромна по обем информация и работата с нея.

Компютърните подходи за идентифициране на важно съдържание попадат в двете крайности:

1. От една страна, темата и важната информация за даден домейн са предварително дефинирани като шаблони – фрейми за извличане на информация, определени предварително от експерти. Наричат ги системи за съвпадение на модели. Ранните систе-

ми за изкуствен интелект разчитаха на техники за съвпадение на модели, за да отговорят на въпроси.

2. От друга страна, традиционните системи работят само с индикатори, идващи от входните данни. Наричат ги синтактични системи. При тях потребителският въпрос се анализира и разделя до синтактично дърво. Използвайки граматични правила, системите разбират структурата на въпросителното изречение.

Системите въпрос – отговор (QAS) дават отговори на въпросите, зададени на естествен език. Подобно на системите, които обобщават текст от документи, QAS разчитат на сложни техники за естественоезиков анализ. Техниките в QAS можем да класифицираме в три категории:

1. Подход за извличане на информация – базиран на извличане на информация. Най-напред обработват въпроса, определят подробности като тип и фокус на въпроса, тип на отговора и след това използват тази информация за създаване на заявката.

2. Подход, основан на знанието – генерира семантично описание на въпроса, който след това се използва за търсене в структурираните ресурси. Подходящ за ограничени домейни като туризма, медицината и транспорта, където не съществуват големи обеми предварително записани документи. Подходите, базирани на знания, се основават на това, че не трябва да се добавят нови правила за всеки нов домейн или език на приложението. Siri на Apple е такъв вид система.

3. Хибриден подход – в хибридните системи за въпрос/отговор, подобно на Watson на IBM, докато въпросът се анализира семантично, отговорите се генерират, като се използват методите за извличане на информация.

История и преглед на системите за изкуствен интелект

Системите за изкуствен интелект са претърпели значително развитие в последните 50 години, но интересът към темата не намалява.

Изследователите на изкуствения интелект обучават невронни мрежи, за да разпознават модели, от средата на 60-те години на миналия век и разработват системи, базирани на знанието, способни да извършват многоетапно логическо мислене и извличане

на изводи, още от началото на 70-те години. Първите са се наричали перцептрони, а последните се наричат експертни системи. Първоначално тези парадигми са изглеждали обещаващо, но след това са се сблъскали с огромни пречки, които са забавили напредъка им с цели десетилетия. Няколко неща са се променили през последните 50 години, което прави преглеждането и използването на съществуващите подходи икономически изгодно.

В случая на така наречените перцептрони или невронни мрежи двете важни допълнителни „липсващи връзки“, необходими да направят машинното обучение търговски практично, са достъпът до достатъчно данни и евтините процесори. Днес, благодарение на интернет, cloud computing и технологиите Big Data¹ се прилагат повсеместно, а нарастващият глобален пазар за игри донесе бързи, евтини, масови процесори.

Необходимостта от стабилен и приложим уеб ИЕЕ стана по-остра в последните години, тъй като обемът на информацията в интернет нараства постоянно и безкрайно и все повече хора получават достъп до данни, съхранявани в различни формални хранилища чрез уеббраузъри.

В тази статия представяме съществуващи системи за изкуствен интелект. Те освен анализатори на текст са и софтуер за машинно обучение с поведенчески алгоритми, които се адаптират към изискванията на потребителя. Въз основа на огромни бази от данни тези системи подобряват полезността си. По-популярни приложения са:

1. **LUNAR** е един от първите NLIDB, създаден през 60-те години като интерфейс към база данни на лунните скали. Системата, разработена от NASA, е предназначена да отговаря на писмено зададени въпроси за геоложкия анализ на лунните скали, върнати от Луната. Използвани са две бази данни – химични анализи и литературни справки. Програмата използва анализатор на разширени мрежи за преход и процедурна семантика на Уудс. Сис-

¹ Big Data – съвкупност от структурирани и неструктурирани данни с големи обеми, които се обработват с голяма скорост и имат различни форми.

темата е била неофициално демонстрирана на Втората годишна научна конференция, посветена на Луната² през 1971 г.

2. **LADDER** е една от първите добри системи за обработка на естествен език. Системата е проектирана като естественоезиков интерфейс към база данни с информация за корабите на американския флот. Тази система, описана от Г. Хендрикс и кол. (Hendrix, Sacerdoti, Sagalowicz, Slocum/Хендрикс, Сасердоти, Сегалович, Слоукъм 1978), използва семантична граматика за разбор на въпроси и запитвания към ограничена база данни. Системата LADDER може да поддържа само прости заявки от една таблица или запитвания към няколко таблици с лесни условия за съответствие.

3. Известна система от средата на 80-те години е **CHAT-80**, която е изпълнена изцяло на Prolog. Целта и усилията са били насочени към въпросите на преносимостта и разработването на общи интерфейси, които могат лесно да бъдат адаптирани към различни бази данни.

4. **SIRI** – личен асистент на Apple с женски глас. Чрез нея намираме информация, изпращаме съобщения, добавяме събития в календарите си, получаваме напътствия. Приложението използва технология за машинно обучение, за да стане „по-умно“ и способно да предвиди и разбере нашите въпроси и искания на естествен език.

5. **Google Assistant** е мултиплатформена услуга, която работи на всякакви устройства. Google Assistant може да открива желаните от потребителя филми и след това да поръча билети за тях. Google Assistant не е обвързан с конкретен смартфон, таблет или друго устройство. След като му се зададе въпрос, той определя контекста, следи разговора и е в състояние да върне коректен отговор.

6. **ALEXA** е въведена за първи път от Amazon. Има невероятната способност да дешифрира речта от всяка точка на стаята, може да ни помогне да пазаруваме, да насрочваме срещи, да настроим аларми, да управляваме интелигентно жилището си, помага на хора с ограничена подвижност.

² Second Annual Scientific Lunar Conference in Houston, 1971.

7. **COGNITO** – стратегия за обслужване на клиенти, която предлага персонализирана поддръжка. Приложението може да разбира контекста, да разговаря с клиентите си на естествен език и да се учи от поведението си. Насочва ги към подходящия продукт въз основа на данни от потребителски профил или събития от живота (например покупка на жилище, получаване на студентски кредит и т.н.). Тези високоцелени, навременни оферти увеличават реализациите и имат положително въздействие както върху компаниите, така и върху клиентите.

8. **BOXEVER** – софтуерно приложение, което чрез използване на изкуствен интелект и машинно обучение във вид на игра помага на потребителите си да намерят нови начини за резервиране на своите пътувания. Целта му е да се подобри взаимодействието между туристическата индустрия и нейните клиенти.

9. **ELIZA** – компютърна програма за естествен език, която симулира разговор, използвайки методология за „съвпадение на модели“ и заместване, която дава на потребителите илюзия за разбиране от страна на програмата, но няма изградена рамка за контекстуализиране на събития (вж. Jurafsky, Martin/Джурафски, Мартин 2008).

10. Диалогова система **HAM-ANS** (Hoeppner, Morik, Marburger/Хопнер, Морик, Марбургер 1984), чието поведение в диалоговия прозорец може да бъде прехвърлено от режим на сътрудничество (например системата отговаря на въпроси) до режим „базиран на интереси“ (например системата се опитва да убеди потребителя да резервира стая в конкретен хотел).

11. Системите **NLMenu** и **NATURALLINK** предоставят възможности за динамично формиране на т.нар. естественоезикови менюта, като по този начин се избягват редица лингвистични трудности. Потребителят може да избира сред голям брой алтернативи, но езикът за въвеждане се контролира. Всеки избор има добре дефинирана синтактична структура. Тъй като качеството на системата за обработка на естествения език е силно зависимо от формалното качество на входа, то тази система се разглежда като идеалния компромис за въздействието/ефекта човек – компютър. Потребителят се ръководи от набор от динамично генерирани

менюта, за да използва подмножеството на естествения език, който разбира.

12. **Language Craft** е инструмент за изграждане на интерфейси на естествения език към интерактивни системи, включително бази данни и експертни системи.

13. **Penn Tree Bank** е електронен ресурс от синтактично аотирани текстове за българския език, който може да бъде:

1. Основа на формална компютърна граматика на българския език (parser);

2. Консистентно аотиран, за да се използва лесно и удобно;

3. Достатъчно детайлно аотиран, за да позволява поспецифични употреби (система за отговаряне на въпроси, подкрепена от семантични графи);

14. **FoG** – система, използвана от Канада, за генериране на прогнози за времето на френски и английски език в началото на 90-те.

15. **Phrasetech, Arria NLG, Automated Insights, Narrative Science, Retresco и Yseop** – преобразуват данните в разкази.

16. **FriendlyData API** – интерфейс за търсене на естествен език. Използва се от фирмите за пазарен анализ, за продажби на дистрибутори и прогнози на продажбите, за проучване на нивата на запаси и веригата на доставки, за пестене на оперативни разходи – с цел да бъдат намерени скрити послания, да се спести време и да се вземат по-бързи бизнес решения.

17. **Sophia Semantic Search** – позволява на системата да разпознае значението на текст (като поредица от символи) и да идентифицира семантични елементи като понятия, обекти и отношения.

18. **CLIR** – извличането на информация на различни езици (CLIR) е подобласт на информационното извличане, която се занимава с извличане на информация, написана на език, различен от езика на заявката на потребителя. Целта на CLIR е да намери информацията, написана на език, различен от езика на заявката.

19. **LORELEI** (Low Resource Languages for Emerging Incidents) е програма за инциденти. Езиковите пакети се състоят от големи обеми от формален и неформален едноезичен и паралелен (с английски) текст с разнообразни ръчни анотации за подкрепа на ситуационната осведоменост, плюс лексикон и основни инстру-

менти за обработка. Езиковите пакети за инциденти съдържат ръчно обозначени данни за оценка, предназначени да тестват представянето на системата за задачи, свързани със ситуационната осведоменост за един или повече езици.

20. The **DARPA** Machine Reading Program – прави информацията, съдържаща се в текста, достъпна във форми, които компютрите могат да използват за автоматизирана обработка. Изградени са текстови модели, които обхващат знанията от естествено въведен текст и го трансформират във формални представяния, използвани от изкуствения интелект.

21. **Hydra** – българската система Hydra е платформенонезависима програма за създаване и валидиране на лексикално-семантични мрежи. Тя дава възможност на потребителите да редактират или търсят в произволен брой уърднети едновременно. Отделните уърднети могат да се синхронизират, което позволява едновременното визуализиране на еквивалентните синонимни множества в различните езици.

22. **Chooser** е многофункционална платформенонезависима система, която позволява дефиниране на различни анотационни схеми. Основните функции на програмата са: бързо и интуитивно анотиране; достъп до детайлна информация за предложените значения чрез синонимните множества и информацията, свързана с тях – синоними, тълковни дефиниции, семантични релации; идентифициране на съставни думи и изрази както с елементи в съседство, така и с дистантни елементи; разнообразни стратегии за обхождане на корпуса; операции върху езиковите единици – редактиране, вмъкване, изтриване на думи и символи; гъвкава стратегия за търсене в корпуса, чрез прости заявки и регулярни изрази по произволна или основна форма на дума (от официалния сайт на Chooser).

23. Проектът **FrameNet** на Университета в Бъркли се базира на теория на значенията, наречена фреймова семантика, в чиято основа лежи схващането, че значението на повечето думи може да бъде разбрано най-добре на базата на семантичните фреймове (описание на типа събитие, отношение, процес или явление) и на участниците в тях. Лексикалната база данни FrameNet е едновременно четима за човека и машината въз основа на анотиране на

примери за това – как се използват думите в действителните текстове. Съдържа предварително фиксиран брой текстови единици с предварително зададена структура – повече от 200 000 изречения, свързани с повече от 1200 семантични рамки, които осигуряват уникален набор от данни за обучение за семантично етикетирание на ролята, използвани в приложения като извличане на информация, машинен превод, разпознаване на събития, анализ на чувствата и др.

24. **WordNet** е система, разработена в Университета в Принстън като лексикална база данни за английски език. WordNet според П. Несторова (Несторова/Nestorova 2008) може да се нарече многофункционален речник – тълковен, синонимен – и освен това поради йерархичността в подреждане на лексемите дава възможност да се проследи в каква система влиза отделно понятие. WordNet е форма на традиционна семантична мрежа. Възлите в нея са множества от действителни думи в английския език, които имат общо значение и са синоними в даден контекст. Дъгите изразяват семантични, морфо-семантични и екстралингвистични релации между възлите (от официалния сайт на BulNet).

25. **VerbNet** е проект, в който предикатите са съотнесени както с фреймовете във FrameNet, така и с понятията в WordNet (Schuler/Шулер 2005).

26. **BulNet** – част от многоезична семантична мрежа на балканските езици BalkaNet. Българският WordNet съдържа повече от 80 000 синонимни множества, разпределени в девет части на речта – съществителни, глаголи, прилагателни, наречия, местоимения, предлози, съюзи, частици и междуметия. Изграден е от множества от думи, наречени синсети (synsets – synonymy sets). Всяко синонимно множество – synset, кодира релация на еквивалентност между няколко единици – literals (поне един literal трябва да присъства експлицитно в множеството), които имат уникално лексемно значение, принадлежат към една и съща част на речта и изразяват еднакво значение.

27. **Семинвест** – общото между FrameNet и Семинвест е в представянето на семантиката на лексикалните единици чрез набор от свързани помежду си понятия, т.е. чрез концептуални структури. Във FrameNet не само глаголите, но и съществителни-

те, прилагателните, наречията и предлозите имат семантични рамки (semantic frames), а в Семинвест са включени само глаголни единици.

28. **ATLAS** – Applied Technology for Language-Aided CMS – проект, насочен към създаването на система за публикуване и поддръжка на електронно съдържание на различни езици. Неговата основна цел е да улесни разработването и управлението на многоезичното уебсъдържание, по-специално създаването и поддръжката на многоезични уебсайтове (от сайта на проекта). Системата включва съществуващи програмни ресурси за компютърна обработка на български, гръцки, румънски, хърватски, полски, немски и английски. Системата има три компонента: с **iLibrarian** и **iPublisher** потребителите могат лесно да създават, поддържат и публикуват многоезиково съдържание в интернет, а с **EUDocLib** имат достъп до европейските правни документи. **iPublisher** е **SAS** услуга за уебсъдържание, която позволява на малки и големи организации да внедряват и управляват многоезични уеб сайтове, без да харчат време и усилия за инсталиране и поддържане на система за управление на съдържанието. **iLibrarian** е уеббазирана услуга за съдържание, която позволява на потребителите не само да съхраняват, организират и публикуват лични творби, но и да намират подобни документи на английски и български и лесно да получават най-важните текстове от големи колекции от непознати документи или резултати от търсачката.

Основните услуги, които се предлагат, са: автоматична категоризация на документи, автоматично генериране на резюмета на документи, автоматичен превод на резюметата, извличане на думи и изрази, които са „най-важни“ за даден документ, интелигентно търсене в многоезичното съдържание.

29. **UCREL Semantic Analysis System** е софтуерен инструмент за извършване на автоматичен семантичен анализ на устни и писмени данни на английски език. Лексикалните ресурси, на които се основава системата, се конструират ръчно, а също така съществуват методи за определяне на недвусмисленост, включително маркиране на част от речта, класиране за обща вероятност, извличане на много думи, област на идентификация на дискурса и контекстуални правила.

30. **НетПеак** – с помощта на този инструмент думите в текста могат да се сведат до основните им форми и да се види какво е семантичното ядро на текста. Като допълнение има възможност и за търсене на синоними, общи статистики на текста.

31. Платформата **Ellogon** управлява и пази текстови данни, създава и управлява езикови модели за обработка и накрая представя текстовите данни и свързаната с тях информация.

32. Системата **GATE** може да извлича информация, да анализира текст.

33. Системата **PLINIUS** е предназначена за автоматично извличане на знания от текстове в областта на химията. При нея са определени няколко множества от понятия (като химичен елемент, цяло число, група и др.) и правила за конструиране на останалите понятия. В онтологията, описваща системата, има представени около 150 понятия и 6 правила (вж. Терзиева, Тотков/Terzieva, Totkov 2009). Описва се с помощта на фреймови конструкции.

34. **СУС** на **Сусорг** е система, ориентирана към създаването на огромна мултиконтекстна база от знания и специална машина за извод. Основната цел на проекта е построяването на база от знания за всички общи понятия, която включва семантичната структура на термините, връзките между тях и множество от аксиоми, разгледани в работата на Т. Терзиева и Г. Тотков (Терзиева, Тотков/Terzieva, Totkov 2009). Докато машинното обучение използва **big data**, за да прави статистически разсъждения, то **СУС** използва знанията, за да направи каузална аргументация. Базата знания на **СУС** включва: онтология от около 1,5 милиона понятия; повече от 20 милиона общи правила и твърдения, включващи тези понятия; специфични за предметната област разширения на онтологията и базата от знания.

35. Системата **Generalized Upper Model** (подробно за нея вж. Gomez-Perez, Fernando-Lopez, Corcho/Гомез-Перез, Фернандо-Лопез, Корчо 2004), е ориентирана към поддръжка на процесите на обработка на естествен език (английски, немски и италиански). Равнището на абстракция се намира между лексическите и концептуалните знания и се определя от изискванията за опростяване на интерфейсите с лингвистични ресурси. Моделът на **Generalized**

Upper Model включва таксономия, организирана под формата на йерархия между категориите (около 250 понятия), и отделна йерархия на връзките.

36. Софтуерът **KUERI** интегрира текстово поле за търсене в релационни SQL или NoSQL бази данни. Крайните потребители могат да зададат въпроси на естествения си език и Kueri ще преобразува тези въпроси в правилно форматирани SQL или JSON заявки за бази данни. Също така може да предостави предложения, които помагат на потребителите при конструирането на въпросите. Kueri базира своето предложение върху свързаната структура на базата данни, върху части от стойностите и съдържанието на базата данни и върху историята на въпросите на потребителя.

Предизвикателства и бъдеща работа

Системите с естественоезиков интерфейс към база данни са изпълнени с редица проблеми и предизвикателства, които обясняват и големия брой софтуерни продукти в тази област.

Проблем с „празния ред“ (известен още като проблем за обитаемост). Всъщност в основата си е тревожно да се намериш пред празен ред. Потребителите се нуждаят от някакви насоки за това – от къде да започнат, какво да питат, и трябва да разберат какво е възможно да се зададе като въпрос.

Друг проблем е, когато системата не разбере въпрос – не е ясно на потребителя дали въпросът е отхвърлен заради езиково покритие, или е извън концептуалното покритие на системата.

Естественоезиковата генерация е задача за генериране на заключения – разбираеми отговори, текстове на естествен език от компютърно софтуерно приложение, например база данни или логическа форма, каквато е фреймът. Фрейми са и източниците, от които се извличат данните в голяма част от разгледаните системи. Като бъдеща задача ни предстои да представим модел за класифициране на понятия, релации и правила в ограничена предметна област, структурно организирани в йерархия от фрейми. Ще се добави и синтактико-семантичен анализатор. Моделът ще се изследва експериментално и ще бъде анализиран в последващи публикации. Естественоезиковият интерфейс за диалог със системата ще се осъществи като начало на български език.

В компютърните системи за изкуствен интелект представянето на базата знания е близко до начина, по който се съхраняват понятията в човешкия мозък.

Обработката на естествения език до заявка към базата данни е вид преобразуване. Естественоезиковата генерация, т.е. от базата данни към текст на естествен език, е също преобразуване, но се използват различни методи.

Изследванията сочат, че текстовите резюмета могат да бъдат по-ефективни от таблици, графики и други визуални образи за вземане на решения и че компютърно създадените текстове могат да бъдат по-добри (от читателска гледна точка).

Технологиите, свързани с естественоезиковата обработка, зависят в значителна степен от езиковите ресурси и средствата, които могат да бъдат използвани.

Следователно работата в насока обогатяване на корпусите с анотирани ресурси и инструменти, описващи синтактично и семантично езиковата инфраструктура, е съществена за развитието и внедряването на системи с естественоезиков интерфейс към бази данни и в български информационни платформи.

БИБЛИОГРАФИЯ

ATLAS project/АТЛАС проект: <<http://www.atlasproject.eu/>>.

BulNet/БулНет: <<https://dcl.bas.bg/bulnet/>>.

Gomez-Perez, Fernando-Lopez, Corcho/Гомез-Перез, Фернандо-Лопез, Корчо 2004: Gomez-Perez, A., M. Fernando-Lopez, O. Corcho. *Ontological Engineering with examples from the areas of Knowledge management, e-Commerce and the Semantic Web*. London: Springer-Verlag – 2nd. Ed., 2004.

Hoerppner, Morik, Marburger/Хопнер, Морик, Марбургер 1984: Hoerppner, W., K. Morik, H. Marburger. Talking it Over: The Natural Dialog System HAM-ANS. // *Cooperative Interfaces to Information Systems*. Berlin: Springer, 1984, 189 – 258.

Hendrix, Sacerdoti, Sagalowicz, Slocum/Хендрикс, Сасердоти, Сеголович, Слоукъм 1978: Hendrix, G., E. Sacerdoti, D. Sagalowicz, J. Slocum. Developing a natural language interface to complex data. // *ACM Transactions on Database Systems*, 3(2), June 1978, New York: SRI International, 105 – 147.

Jurafsky, Martin/Джурафски, Мартин 2008: Jurafsky, D., J. H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Second edition, New Jersey: Prentice Hall, 2008.

Несторова/Nestorova 2008: Несторова, П. *Семантични и синтактични особености на предикати в българския език*. Пловдив: Весела, 2008. [Nestorova,

P. *Semantichni i sintaktichni osobenosti na predikati v balgarskiya ezik*. Plovdiv: Vesela, 2008.]

Rukshan, Prashanthi, Sinnathamby/Рукшън, Прашантн, Синатамби 2013: Rukshan, A., R. Prashanthi, M. Sinnathamby. *Natural Language Web Interface for Database (NLWIDB)*. Proceedings of the Third International Symposium, SEUSL: Oluvil, Sri Lanka, 6 – 7 July 2013.

Терзиева, Тотков/Terzieva, Totkov 2009: Терзиева, Т., Г. Тотков. Един онтологичен подход към процеса на обучение. // *Сборник доклади на Национална конференция „Образованието в информационното общество“* Пловдив, АРИО, 12 – 13 май 2009, 17 – 24. [Terzieva, T., G. Totkov. Edin ontologichen podhod kam protsesa na obuchenie. // *Sbornik dokladi na Natsionalna konferentsiya “Образованието в информatsionното obshtestvo”* Plovdiv, ARIО, 12 – 13 may 2009, 17 – 24.]

WordNet/УърдНет: <<https://wordnet.princeton.edu/>>

FrameNet/ФреймНет: <<https://framenet.icsi.berkeley.edu/fndrupal/frameIndex>>

Schuler/Шулер 2005: Schuler, K. *VerbNet: A broad-coverage, comprehensive verb lexicon*, PhD dissertation, University of Pennsylvania, Philadelphia, 2005.