

Автоматично разпознаване на концептуално-семантични отношения в електронен корпус от биологически текстове и приложението му в лексикографската практика

Сия Колковска

Институт за български език „Проф. Л. Андрейчин“ – БАН

Automatic Recognition of Conceptual Relations in a Biological Corpus and its Application in the Lexicography Work

The results of automatic recognition of two basic conceptual relations – hyponymy relation and part-whole relation between terms in a biological corpus are presented in the first part of the paper. The application of automatically extracted hyponym-hyperonym groups and part-whole groups of terms in the process of dictionary making and editing is discussed in the second part. A technique for finding errors and omissions in the presentation of hyponymy relations and part-whole relations between terms in terminological dictionaries is proposed in the paper.

Настоящата работа представлява продължение на предишно изследване върху автоматичното разпознаване на концептуално-семантични отношения (по-нататък КСО) в електронни текстове от химическата област (вж. Колковска 2005). Първата задача, която си поставяме тук, е да проверим ефективността на използвания в първата работа метод за автоматично разпознаване, като го приложим върху текстове от друга научна област и същевременно – да направим необходимите промени и допълнения в този метод с оглед на неговото усъвършенстване. Поради малкия обем на тази работа ще представим резултатите от автоматичното разпознаване на двете основни КСО: род – вид и част – цяло.

Втората ни задача е да покажем по какъв начин резултатите от автоматичното разпознаване на КСО може да намерят приложение в лексикографията и терминографията. Във втората част на статията ще разглеждаме една от възможностите за използване на автоматично открити КСО в лексикографската практика, свързана с коригиране на пропуски и неточности във вече създадени речници.

I. Автоматично разпознаване на концептуално-семантични отношения

1.1. Същност, методи и използван в работата метод

Както е известно, автоматичното разпознаване на отношения в електронни текстове се състои в това изразяващите тези отношения думи

и изрази да бъдат открити автоматично с определен метод, без да е необходим предварителен анализ на текстовете. Автоматичното разпознаване на отношения и в по-широк план – на езикови единици в електронни текстове е много перспективно и бързо развиващо се направление в компютърната обработка на естествения език. То представлява езикова технология с широко приложение при автоматичното извличане на информация от електронни текстове. Перспективността на тази задача се определя и от това, че тя може да бъде използвана като първи етап на една по-обща цел – автоматичното разпознаване на термини в специални текстове.

В цитираната по-горе предишна работа върху автоматично разпознаване на КСО тази задача е осъществена с помощта на лингвистични методи (а не напр. на статистически, основаващи се най-общо на отчитане на честотата на лексикалните единици в текста).

Както е известно, лингвистичните методи се базират на правила, откъдето идва названието им базирани на правила методи (rule-based methods)¹. Тези правила от своя страна се основават на модели, които се прилагат в електронен текст. При съвпадение между прилагания модел и определена дума или израз в текста те се маркират като поредна поява на търсената езикова единица.

Моделите, използвани при лингвистичните методи, са различни по състав – морфологични, синтактични, лексикални или хибридни – лексико-синтактични. Подобно на редица учени, използващи лексикални модели за разпознаване на КСО в специален текст², ние също си служим с този тип модели като най-подходящ с оглед на нашите задачи. Основен компонент на лексикалните модели са думи или съчетания, наречени маркери, представящи онези лексикалните единици в текста, за които се предполага, че изразяват дадено КСО.

Лексикалните модели, създадени в предишното ни изследване за целите на автоматичното разпознаване на КСО, включват два задължителни компонента, а някои от тях – и трети, незадължителен компонент.

1. Маркер – основен компонент, който, както се каза, представя думите и съчетанията, изразяващи съответното отношение в текста. Т. напр. в модела за партитивно отношение „**Х съдържа У**“ маркер е глаголът **съдържа**м (във формата му за 3 л. ед. ч.). Моделът „**У е компонент на Х**“ включва поредица от три маркера: гл. **съм** в 3 л. ед. ч. + същ. име **компонент** + предлог **на**. Както се вижда от тези примери, някои модели включват само един маркер, а други – поредица от маркери.

¹ Лингвистични методи за автоматично разпознаване на КСО използват Дейвидсън (Дейвидсън 1998), Кондамин (Кондамин 1998), Херст (Херст 1992), Л'Ом, Фелиу (Фелиу 2000а, Фелиу 2000б).

² Лексикални модели използва напр. Фелиу (Фелиу 2002а, Фелиу 2002б) при разпознаването на КСО в специални текстове на испански език.

2. Елементи X и Y, с които се представят условно термините, участващи в отношението, изразено в един модел. Разположени са в началото и края на моделите (в типичния случай).

3. Лингвистична променлива – незадължителен компонент, представящ тези лексикални единици в текста, които са разположени между изразяващите дадено отношение думи. Т. напр. моделът „**У е [прил.] компонент на X**“ включва променлива **{прил.}**, която е разположена между маркерите **е** и **компонент** и представя прилагателните имена, които се явяват в текста в посочената позиция. С него се разпознава поява на отношението част – цяло в изречение 1):

1) *Хумусът <ОЦЦ> е важен компонент на </ОЦЦ>³ почвата.*

Към много от основните модели са създадени вариантни (вторични) модели, получени от основните модели чрез преобразувания от морфологичен и/или синтактичен характер. Т. напр. към модела за КСО част – цяло „**У е градивна единица на X**“ чрез промяна в словореда е конструиран вторичен модел „**Градивна единица на X е У**“. Създаването на вторични модели е продиктувано от стремежа да се обхванат възможно най-много различни типове езикови изкази на отделните КСО в специален текст. Този стремеж е свързан с желанието да разпознаем максимален брой появи на КСО. Така например чрез посочения по-горе вариантен модел „**Градивна единица на X е У**“ се разпознава поява на отношението част – цяло в изречение 2):

2) *<ОЦЦ> Градивна единица на въглехидратите е </ОЦЦ> монозахаридът.*

Тази поява не би могла да бъде открита с основния модел „**У е градивна единица на X**“.

1.2. Автоматичното разпознаване на концептуално-семантични отношения в електронен корпус от биологически текстове – изводи и резултати

Както се каза, първата задача, която си поставяме тук, е да установим приложимостта и ефективността на изходните модели, конструирани за целите на автоматичното разпознаване на КСО в химическата област, в друга естественонаучна област – биологическата. Един от очакваните резултати от настоящата работа е свързан с обогатяване на изходния списък от модели въз основа на направените наблюдения върху тяхната продуктивност в биологическата област.

Предварителен етап, предшествващ същинската работа, е създаването на електронен корпус от биологически текстове (чрез сканиране на учебна литература), който включва около 300 000 думи.

³ Таговете <ОЦЦ> и </ОЦЦ> (съответно отварящ и затварящ таг), с които са маркирани изразителите на търсеното отношение, означават „отношение част – цяло“.

При прилагане на изходните модели на КСО част – цяло и род – вид за автоматично разпознаване на тези отношения в биологически текстове се стигна до следните изводи:

1.2.1. Установи се, че тези модели дават много добри резултати при разпознаването на посочените две отношения в биологическата област. Това потвърждава предварителното ни очакване те да бъдат приложими и извън химическата област, за която бяха конструирани първоначално. С тях се постигат относително високи стойности на показателите **точност** (presicion) и **пълнота** (recall)⁴ – показатели, които, както е известно, дават обективна представа за ефективността на съответния метод за автоматично разпознаване. При разпознаване на КСО част – цяло и род – вид в биологически текстове с помощта на изходните модели се постигна **точност 93.2%** и **пълнота 72.4%**. Тези стойности са сравними с резултатите, получени при прилагане на моделите в текстове от химическата област – съответно **точност 94.7 %** и **пълнота 80.3 %**. Получените по-ниски стойности на показателя пълнота се дължат на това, че известна част от появите на разпознаваните отношения в биологически текстове имат специфичен езиков израз, който не е отразен в изходните модели.

1.2.2. Установи се, че има известна, макар и не особено съществена разлика в продуктивността на отделните модели в биологическата и в химическата област. Под продуктивност разбираме броя на появите на търсеното отношение, които са разпознати със съответния модел.

Като цяло повечето модели с по-висока продуктивност в химическата област са продуктивни и в биологическата – т.е. с тях се разпознават относително голям брой появи на търсените отношения в текста. Такива са напр. моделите за партитивни КСО „**Х съдържа У**“ и „**У изгражда Х**“ (особено вариантния модел „**Х е изграден от У**“)⁵. Висока продуктивност и в двете области имат и следните три модела за родово-видови отношения:

„**У е Х, предлог? който / чийто / причастие**“⁶;

„**Х бива следните видове: У₁...У_n**“ с вариант „**Х бива У₁...У_n**“;

⁴ С показателя пълнота се установява какъв процент от всички появи на търсените лексикални единици в един текст са разпознати чрез използвания метод. С показателя точност се установява какъв процент от разпознатите с даден метод думи и изрази действително са появи на търсените лексикални единици.

⁵ С цел по-лесно възприемане на моделите тук ги представяме в по-съкратен вариант, в който маркерите присъстват само с формата си за ед ч. В пълния им вид, в който ги прилагаме в текстовете, маркерите присъстват и с формата си за мн. ч. Т. напр. пълният вид на модела „**Х е изграден от У**“ е „**Х е изграден / са изградени \ от У**“.

⁶ Знакът / се използва за изразяване на дизюнкция и отделя два маркера, които са алтернативни или техни форми. Използваният стандартен символ ? означава, както е известно, нула или един елемент, т.е. думата, след която той се намира, може да липсва в текста или да се яви в него един път.

„У се нарича / наричаме X, предлог? който / чийто / причастие“.

Вж. напр. разпознатите с моделите „X е изграден от У“ и „X бива У₁...У_n“ появи съответно на партитивно и на родово-видово отношение между биологически термини в изречения 3 и 4):

3) *Вирусите <ОЧИ> са изградени от </ОЧИ> нуклеинови киселини и белтъци.*

4) *В зависимост от това дали са изградени от еднакви или различни мономери, биополимерите <ОРВ> биват </ОРВ>⁷ хомополимери и хетерополимери.*

Същевременно се установи, че някои от изходните модели проявяват висока продуктивност в биологическата област, която не им е присъща в химическата област като напр. моделите за КСО част – цяло „X включва У“, „X е съвкупност от У“ и „У е част от X“. С по-ниска честота в химическите текстове е и продуктивният в биологическата област модел на родово-видово отношение „X се разделят на числ.? прил.? вида / групи: У₁...У_n“ и неговият вариантен модел „X се разделят на У₁...У_n“. Като примери ще посочим разпознатата с модела „X се разделят на числ.? прил.? вида / групи: У₁...У_n“ поява на КСО род – вид между биологически термини в изречение 5):

5) *В зависимост от това в коя част на денонощието са активни, животните <ОРВ> се разделят на три групи: </ОРВ> дневни, нощни и сумрачни животни.*

Наблюденията върху продуктивността на изходните модели в биологическата област са полезни, тъй като ни дават информация за това кои от тях са с по-широка приложимост и може да разчитаме на резултатността им в повече научни области.

1.2.3. Установи се, че има известна разлика и в степента на надеждност на моделите в двете области. При прилагането на някои изходни модели в биологически текстове се наблюдава по-висок процент на грешно разпознати думи и изрази в текста (т.нар. шум). Такива случаи бяха установени при два модела с висока продуктивност в биологическата област, а именно – при модела за партитивно КСО „X изгражда У“ и при модела за родово-видово КСО „X се дели на У₁...У_n“ (вариантен модел на първичния модел „X се дели на видовете / групите У₁...У_n“. Вж. един пример за грешно разпознаване чрез втория модел в изречение б), в което реално няма поява на търсеното отношение:

б) *При размножаването клетката <ОРВ> се дели на </ОРВ> две части. – грешно разпознаване⁸.*

⁷ Таговете <ОРВ> и </ОРВ> (съответно отварящ и затварящ таг), с които са маркирани изразителите на търсеното отношение, означават „отношение род – вид“.

⁸ Сравни с вярно разпознатата с този модел поява на родово-видово отношение в изречение 9): *В зависимост от произхода им биоценозите <ОРВ> се делят на </ОРВ> естествени и изкуствени биоценози.* Наличието на грешно разпознати появи с модела с маркер **деля се** е следствие от неговата многозначност като лексикална единица.

Направените наблюдения върху степента на надеждност на отделните модели в биологическата област са важни с оглед на установяване с по-голяма точност на това кои от тях са рискови и трябва да се прилагат по-внимателно, със задължително валидиране на разпознатите появи.

1.2.4. Най-ценният резултат от настоящата работа – това е конструирането на нови модели на родово-видовите и партитивните КСО, чрез което се обогатява изходният списък от модели.

Новите модели се създават по емпиричен път, чрез обобщаване на най-типичните пропуски при автоматичното разпознаване на КСО в биологически текстове с помощта на изходните модели. Появите на отношенията род – вид и част – цяло, които остават неразпознати, се установяват при валидирането на резултатите от автоматичното разпознаване. Валидирането се извършва чрез ръчно преглеждане на текстовете, върху които е приложено автоматично разпознаване, като целта му е да се установяват случаите на грешно разпознаване и на пропуснати появи на търсените отношения.

Въз основа на наблюденията, направени при валидирането на резултатите от автоматичното разпознаване с помощта на изходните модели, се конструират още 9 модела на КСО род – вид и част – цяло. Тук ще посочим само някои от тях.

Един от най-типичните случаи на неразпознати появи на родово-видови отношения между биологически термини е повод да бъде конструиран моделът „Съществуват числ. прил.? типа / вида $X - Y_1...Y_n$ “. С добавянето на този модел към изходния списък става възможно откриването например на родово-видовото КСО между биологическия термин *клетка* (родов термин) и термините *еукариотна клетка* и *прокариотна клетка* (видови термини) в изречение 7):

7) <ОРВ>Съществуват два основни типа <ОРВ> клетки – еукариотни и прокариотни (доядрени).

Допълнително конструирани са и следните два модела на родово-видови КСО:

„ X са групирани в числ. прил.? в групи $X: Y_1...Y_n$ “;

„Известни са числ. прил.? типа / вида / $X: Y_1...Y_n$ “.

Вж. откритите с тези модели появи на родово-видови отношения между биологически термини съответно в изречения 8) и 9):

8) Почвените обитатели <ОРВ> са групирани в три екологични групи <ОРВ>: геобионти, геофили и геоксени.

9) <ОРВ>Известни са три основни вида <ОРВ> РНК: информационни, транспортни и рибозомни.

Сред новите модели на партитивни КСО е моделът „ $У$ е градивна единица на X “, който има вариантен модел „Градивна единица на X е $У$ “

с висока продуктивност. Вж. разпознатата с вариантния модел поява на търсеното отношение в изречение 10):

10) *Мономерната <ОЧЦ> градивна единица на белтъците е </ОЧЦ> аминокиселината.*

Обогатяването на изходния списък от модели е важен етап в процеса на усъвършенстване на използвания метод за автоматично разпознаване на КСО. По този начин се повишава единият от показателите за неговата ефективност – показателят пълнота, като води до увеличаване броя на откритите появи на търсените отношения в текста.

И така, при прилагане на конструираните от нас модели за разпознаване на КСО част – цяло и род – вид в електронни текстове от биологическата област се потвърди тяхната надеждност и ефективност. Бяха създадени нови модели, чрез което се направи поредна стъпка в процеса на доразвиване на използвания метод за автоматично разпознаване на КСО.

II. Приложение на автоматичното разпознаване на концептуално-семантичните отношения род – вид и част – цяло в лексикографската практика

Актуалността на автоматичното разпознаване на КСО род – вид и част – цяло се обуславя от особената важност на тези отношения за лексикалната система, в частност – за терминологичните подсистеми. Те са основният тип системни връзки между понятията, респ. – между лексикалните единици. КСО род – вид и част – цяло са основа на системната организация на понятията и съответно имат огромно значение (заедно със семантичните отношения синонимия и антонимия) за изграждане на лексикалната система

Важността на КСО се обуславя и от друга причина. В текста те се проявяват между лексикални единици (в специален текст – между термини). Чрез автоматичното разпознаване на тези отношения се откриват автоматично и групите от лексикални единици, свързани със съответното отношение. Т. напр. вследствие на разпозната поява на КСО род – вид в изречението 11):

11) *<ОРВ> Съществуват два основни типа </ОРВ> клетки – еукариотни и прокариотни.*

автоматично се открива следната група от родов термин и видови термини: **родов термин:** *клетка* – **видови термини:** *еукариотна клетка, прокариотна клетка.*

От своя страна, знанието за групите от лексикални единици, свързани с родово-видови или партитивни отношения, е полезно както при изучаването на лексикалната (в частност – на терминологичната) системност,

така и при отразяването на тази системност в речници, бази данни и семантични мрежи. Това знание е особено важно за терминологията поради организирането на термините от различните понятийни области в строги системи, изградени главно върху посочените две отношения.

Важността на КСО род – вид и част – цяло като основен тип системни отношения, от една страна, и от друга страна – на знанието за групите от свързаните чрез тях лексикални единици обуславя редица направления в приложението на автоматичното разпознаване на тези отношения.

Извлечени от корпуси или електронни речници КСО се използват за автоматично построяване на понятийните системи в някои специални области под формата на терминологични бази знания и понятийни онтологии⁹. Те имат приложение и при автоматичното построяване на системата на термините в определени области под формата на терминологични бази данни и езикови онтологии, а също и за автоматично построяване на лексикални йерархии¹⁰.

Списъци с автоматично извлечени групи от свързани с КСО лексикални единици (предимно родово-видови и партитивни групи) се използват при работата върху речници, бази данни и семантични мрежи. Те намират приложение при създаване на тезауруси, а също и за подобряване представянето на системността в съществуващи речници, тезауруси и семантични мрежи. При съпоставянето на тези автоматично извлечени групи от лексикални единици с имплицитно присъстващите във всеки тълковен речник и семантична мрежа системни отношения лесно се установят някои непоследователности и пропуски в тях. Т. напр. Херст оценява пълнотата и точността на семантичната мрежа WordNet, като проверява как са представени в нея автоматично откритите от него групи от хипероними и хипоними (Херст 1992).

Като илюстрация на приложението на автоматично извлечени КСО в лексикографията ще споменем опита си в използването на такива отношения при изработване на речници. В процеса на работата върху частта „Химия“ на „Терминологичен речник по природни науки“ (Попова и др., 2006) използвахме като допълнителен източник на информация родово-видовите и партитивни групи от термини, извлечени от нас автоматично от електронни текстове от химическата област. Това ни помогна в редица случаи за по-точното представяне на системните връзки между химическите термини (вж. Колковска 2008).

⁹ Вж. Дейвидсън 1998, Мейер 1997, Л'Ом, Фелиу 2002б. Фелиу напр. използва автоматично открити КСО за построяване на онтология на знанието в областта генетика (раздел човешки геном).

¹⁰ Т. напр. Херст (Херст 1992) използва автоматично откритите от него родово-видови отношения в електронен корпус за автоматично подреждане в йерархия на свързаните чрез тези отношения думи.

Тук ще илюстрираме полезността на автоматично извлечени родово-видови и партитивни групи за лексикографската работа, като използваме такива групи от термини от друга терминологична подсистема – биологическата. Списъкът с тези групи от термини бе извлечен чрез обобщаване на появите на родово-видови и партитивни КСО, които разпознахме в корпуса от биологически текстове. Тук ще разгледаме как те могат да бъдат използвани за подобряване представянето на системността във вече изработени речници.

Извлечените групи от биологически термини бяха сравнени с начина на представяне на системните връзки между съответните термини в „Речник по биология“ (Маринова 2004). Откритите случаи на несъответствия, след съответни справки¹¹, бяха обобщени и може да послужат при коригиране представянето на системността в този речник в негови следващи издания.

За целите на направената съпоставка бе избран именно посочения речник, тъй като той включва биологически термини, изучавани в училище и следователно сравними с групите от термини, извлечени от нас от учебна литература за средните училища.

При направената съпоставка се установи, че в повечето случаи е налице съответствие между установените от нас родово-видови и партитивни връзки (респ. – групи от биологически термини) и между включените в „Речник по биология“ (по-нататък РБ) такива връзки. Установените случаи на несъответствие са относително малко на брой, което е основание за извода, че представянето на системността в РБ е прецизно и на добро равнище. Използвана тук методика ни дава възможност да оценим обективно това качество на един речник.

Случаите на несъответствия се дължат на недостатъчна пълнота или недостатъчна точност на отразяване на отделни родово-видови и партитивни отношения в разглеждания речник.

1. Недостатъчна пълнота в отразяването на някои родово-видови и партитивни отношения

Този тип несъответствия най-често е свързан с пропуснати родово-видови и партитивни връзки, в които участват включени в речника термини. В по-редки случаи недостатъчната пълнота се изразява в пропускане на отделни членове на родово-видови и партитивни групи.

¹¹ В някои случаи представянето на родово-видовите и партитивните отношения в учебниците не е коректно и може да доведе до погрешно изведени групи от термини. При съмнение в коректността на автоматично изведената група от термини се правят допълнителни справки с друга справочна литература.

1.1. Пропуснати родово-видови и партитивни връзки на термините

Най-често пропускът е свързан с отсъствие на родово-видова или партитивна група на включен в речника термин. Установи се например отсъствието в РБ на една важна в биологическата подсистема родово-видова група: **родов термин:** *организми*, **видови термини:** *прокариоти* и *еукариоти*. Всъщност в посочения речник липсва и самият родов термин *организъм*, въпреки че повечето от родово-видовите групи, в които той участва, са включени.

Установеният пропуск може да бъде използван за обогатяване представянето на системността в РБ чрез включване на пропуснатата концептуално-семантична връзка. Възможността за коригиране представянето на системността в един терминологичен речник в посочената посока е много полезна с оглед на явлението многопосочност на системните отношения (multidimensionality) (вж. Боукър 1997, Кагеура 1997). То се разглежда предимно по отношение на терминологията и е свързано с възможността един термин да участва в повече от една системна връзка от един и същ тип (т. напр. да бъде член на повече от една родово-видова или партитивна група). Напълно е възможно някои от тези групи да бъдат пропуснати при лексикографското описание на дадена терминологична подсистема.

Чрез сравнение с автоматично извлечени КСО откриването на такива пропуски значително се улеснява. Така например чрез съпоставка с извлечените от нас родово-видови и партитивни групи от биологически термини се установи отсъствието в РБ на една от партитивните връзки на термина *клетка*. Като название на цяло той участва в две партитивни групи: **I. термин за цяло:** *клетка* – **термини за части:** *нуклеинови киселини*, *белтъци*, *липиди* и **II. термин за цяло:** *клетка* – **термини за части:** *цитоплазма* и *ядро*. В РБ е отразена само първата партитивна връзка на този термин. Посочената по-горе като липсваща в РБ видова двойка *прокариоти* и *еукариоти* се отнася към родовия термин *организъм*, на който също е присъщо явлението многопосочност.

1.2. Пропуснати членове на включени в речника родово-видови и партитивни групи

Пример за такъв тип непълнота е отсъствието в РБ на единия видов термин в родово-видовата група с **родов термин:** *животни* и **видови термини:** *водни обитатели (хидробионти)* и *почвени обитатели*. Терминът *почвени обитатели* липсва в посочения речник.

2. Недостатъчна точност в отразяването на някои родово-видови и партитивни отношения

В РБ са открити отделни случаи на недостатъчна точност, свързани с тълкуването на група видови термини чрез отнасянето им към прекалено

общ родов термин вместо към непосредствения родов термин. Т. напр. терминът *водни обитатели* (със синоним *хидробионти*) участва в родово-видовата група: **родов термин:** *водни обитатели (хидробионти)* – **видови термини:** *планктон, нектон и бентос*. В определенията на тези видови термини в РБ като родов термин е посочен терминът *организъм*, а не непосредственият родов термин *хидробионти* (макар че в описанието на термина *хидробионти* са посочени видовете му термини). С помощта на използваната методика този и други такива подобни случаи бяха открити сравнително лесно.

Установените случаи на недостатъчна пълнота или недостатъчна точност в отразяване на отделни родово-видови и партитивни отношения в РБ може да бъдат използвани за обогатяване и прецизиране на представянето на системните връзки в речника чрез допълнително включване на пропуснатите термини или коригиране на някои терминологични дефиниции при негово следващо издаване.

И така, във втората част на работа се илюстрира как, като се използват автоматично извлечени от текстове родово-видови и партитивни отношения, може лесно да се установят пропуски в представянето на системността в специални речници. Несъмнено отразяването на системността в един речник може да се усъвършенства и по други начини. Предимство на разглежданата тук методика е, че чрез нея този процес на усъвършенстване значително се улеснява и ускорява. Освен това тя дава възможност и на неспециалисти в съответната научна област да откриват такива непълноти и неточности.

Тук разгледахме по-подробно едно от приложенията на автоматичното разпознаване на КСО в лексикографската практика, а именно – за подобряване представянето на системните връзки във вече създадени речници. Както бе посочено, автоматичното разпознаване на КСО има много по-широки сфери на приложение. Автоматично извлечени групи от термини може да бъдат полезни и в процеса на изработване на специални и тълковни речници на общоупотребимия език, а също и при изграждането на терминологични бази данни.

Цитирана литература

БОУКЪР 1997: L. Bowker. Multidimensional Classification of Concepts and Terms. In: Handbook of Terminology Management. Vol. 1. Basic Concepts of Terminology Management. Eds. S. Wright and G. Budin. John Benjamins Publishing Company, Amsterdam/Philadelphia, 1997, pp. 131–143.

ДЕЙВИДСЪН 1998: L. Davidson, J. Kavanagh, K. Mackintosh, I. Meyer, D. Skuce. Semi-automatic Extraction of Knowledge-rich Contexts from Corpora. In: Proceedings of Computerm '98, 1998, pp. 50–56.

- КАГЕУРА 1997: K. Kageura. Multifaceted/Multidimensional Concept Systems. In: Handbook of Terminology Management. Vol. 1. Basic Concepts of Terminology Management. Eds. S. Wright and G. Budin. John Benjamins Publishing Company, Amsterdam/Philadelphia, 1997, pp. 119–132.
- КОЛКОВСКА 2005: С. Колковска. Модели на концептуално-семантичните отношения между термините в специален (химически) текст с оглед на автоматичното им разпознаване. Електронно издателство „LiterNet“, 2005.
- КОЛКОВСКА 2008: С. Колковска. Модели на концептуално-семантичните отношения част – цяло в специални текстове (от областта на химията). В: Изследвания по фразеология, лексикология и лексикография в памет на проф. д.ф.н. К. Ничева. Академично издателство „Проф. М. Дринов“, София, 2008, с. 415–422.
- КОНДАМИН 1998: L. A. Condamines, J. Rebeyrolle. СТКВ: A Corpus-based Approach to a Terminology Knowledge Base. In: First Workshop on Computational Terminology, Computerm '98, Montreal, 1998, pp. 29–35.
- Л'ОМ: Marie-Claude L'Homme. Using Terminological Definitions to Find Conceptual Relations between Terms. <http://www.olst.unmontreal.ca>.
- МАРИНОВА 2004: М. Маринова. Речник по биология 9–12 кл. „Ипсилон“, Варна, 2004.
- МЕЙЕР 1997: I. Meyer, K. Eck, D. Skuce. Systematic Concept Analysis with a Knowledge-based Approach to Terminology. In: Handbook of Terminology Management. Vol. 1. Basic Concepts of Terminology Management. Eds. S. Wright and G. Budin. John Benjamins Publishing Company, Amsterdam/Philadelphia, 1997, pp. 98–118.
- ПОПОВА И ДР. 2006: М. Попова, С. Колковска, Г. Димитрова, А. Христова, Е. Петкова. Терминологичен речник по природни науки. „Наука и изкуство“, София, 2006.
- ФЕЛИУ 2002a: J. Feliu, M. Teresa Cabré. Conceptual Relations in Specialized Texts. In: Proceedings of the 6th International Conference on Terminology and Knowledge Engineering (ТКЕ '02), Nancy, France, 2002, pp. 45–49.
- ФЕЛИУ 2002b: J. Feliu, J. Vivaldi, M. Teresa Cabré. Towards an Ontology for a Human Genome Knowledge Base. In: Proceedings of the 3rd International Conference on Language Resources and Evaluation (LREC '02), 2002, pp. 1885–1890.
- ХЕРСТ 1992: M. Hearst. Automatic Acquisition of Hyponyms from Large Text Corpora. In: Proceedings of the 14th International Conference of Computational Linguistics (COLING '92), Nantes, France, 1992, pp. 539–545.

Използвани съкращения:

КСО – концептуално-семантични отношения

ОЧЦ – отношение част – цяло

ОРВ – отношение род – вид

РБ – Речник по биология