

ФРЕЙМОВ МОДЕЛ ЗА ПРЕДСТАВЯНЕ НА СЕМАНТИЧНИ РОЛИ И ПРОЦЕСИ ЗА ИЗГРАЖДАНЕ НА ЕСТЕСТВЕНО-ЕЗИКОВ ИНТЕРФЕЙС

Мария Жекова, Георги Тотков

ПУ „Паисий Хилендарски“, гр. Пловдив, ул. „Цар Асен“ 24
m_jekova@uft-plovdiv.bg, totkov@uni-plovdiv.bg

Резюме: В настоящата разработка е представено изследване в областта на компютърната лингвистика относно теоретичен модел за семантично описание на лексикални единици и процеси. Моделът сам по себе си представя знания и отразява концептуалната организация на паметта на човек, както и нейната гъвкавост и нагледност. При създаването на подходящ набор от т. нар. „акумулативни фреймови модели“ е необходимо изграждане на коректна структура на базата данни от лингвистична гледна точка чрез описания на понятия, факти, правила, обекти и отношения между тях, осигуряващи както гъвкавост и по-лесно надграждане на базите от данни, така и ефективно извличане на различни типове информация при обобщаване на резултатите. В представения компютърен модел задачата е да се осигури възможност за синтезиране на текстове в отговор на въпроси на ограничен естествен език (български). Изследването е част от проучване, посветено на приложения на фреймови модели за представяне на знания и процеси в е-обучението.

Ключови думи: естествено-езиков интерфейс, фрейм, фрейм-прототип, фрейм-екземпляр, предметна област, естествено-езикова генерация, въпрос-отговор системи

1. Въведение

В съвременните лингвистични изследвания активно се използват методи на когнитивния анализ. Изследването е свързано с прилагането на обектно-ориентиран подход при представяне на езикови единици, и е в областта на езиковите технологии и компютърната лингвистиката.

Естествено-езиковите интерфейси са специален тип компютърно-човешки интерфейси, където лексикални елементи като глаголи, фрази и клаузи са входни параметри в заявки към базата от данни за извличане на информация под формата на текст. Правилна заявка към базата от данни може да се състави, когато след обработка на потребителската заявка (текст на ограничен български език) се извлекат параметрите и се синтезират отношенията и връзките между тях. Изграждането на естествено езикови интерфейси включва изисква не само създаване на диалогов модул от тип „въпрос-отговор“, но и осигуряване на достъп до данни и текстови структури, проектирани на базата на семантични и синтактични класове и отношения.

Цел на разработката е построяването на лингвистично коректен фреймов модел за представяне на данните в информационна система чрез описание на лексикални единици и процеси в конкретната (за дадената ИС) предметна област (ПО).

Съответният компютърен модел за представяне на данните от информационната система предполага построяване на нова база от данни, „напълнена“ със съдържание, извлечено от различни източници. Коректното представяне на архитектурата на системата е ключ към постигане на добри решения на реални проблеми.

Естествено-езиковата система може да се адаптира към конкретна ПО от потребител в ролята на „модератор“. Настройката се извършва чрез дефиниране на понятията, фактите, правилата, аксиомите, обектите, фреймите, таблиците, релациите и концепциите в ПО. Автоматичната обработка на информацията изисква машинно разбираемо представяне на нейната семантика. Семантичните релации между фреймовите модели играят съществена роля при разбирането и представянето на естествен език. Това означава, че компютърът чете, разбира и маркира думи, фрази, категории, изречения, абзаци и цели документи.

2. Анализ на изследванията в областта до момента

Компютърната лингвистика се стреми към създаването на софтуерни програми и приложения, които с помощта на реализираните до момента репрезентативни модели за представяне на знания улесняват достъпа до огромната по обем информация и улесняват работата с нея.

Системите от тип „въпрос-отговор“ дават отговори на въпроси, зададени на естествен език. Подобно на системи, които обобщават текст от документи, системите „въпрос-отговор“ разчитат на сложни техники за естествено-езиков анализ.

Най-популярната и ефективна репрезентативна форма за декларативно описание и представяне на знания е чрез правила за описание на действия посредством шаблони (rule-based system). Правилата, които са в сила всъщност са вид предикатна логика с добавени компоненти, показващи как информацията в правилата да бъде използвана в разсъжденията и имат силата да представят огромен диапазон от заключения и спецификации на поведението, зависещи от ПО.

Сами по себе си обаче, правилата / шаблоните не са подходящи за дефиниране на понятия и обекти и за описание на връзките между тях. Те не предлагат ефективно гъвкаво средство за представяне. Главният им недостатък е в области, които ефективно се описват с помощта на структури от фрейми. Системи като LOOPS [9] и CENTAUR [10] са показали как двата

подхода могат да си взаимодействат. Тези системи показват как фреймовото описание може да служи като основа на правилата за описание на действия.

Задача на разработката е чрез интегрирането на фреймовите структури и правилата за описание на понятия и процеси да формираме подход за хибридно представяне, който съчетава предимството на двете техники за представяне на компонентите.

В статията, чрез фреймови модели, се предлага начин за структурно описване на обектите, споменати в правилата. Иерархията на фреймите може да се използва и ще подпомага индексирването и организирането на правилата за съответствие. Тази възможност пък ще доведе до по-лесното конструиране на системата от правила както и на контрола - кога и с каква цел да се използват отделни колекции от правила.

3. Фрейм - същност

Според [1], фреймът се използва за представяне на минималната структурирана информация, която дефинира еднотипни явления, събития, ситуации или процеси. Понятието е въведено в изкуствения интелект от М. Мински [2] като средство за представяне на знания. Фреймовите модели се използват за представяне на относително независими, логически обособени единици на знанието с възможност за многократно използване в различни ситуации (вкл. извличане и агрегиране на данни) [7].

Всеки фрейм има уникално име и структура, състояща се от подредени елементи, наречени слотове. Той не може да се сведе само до обективното представяне на описваната ситуация или процес, в него всеки слот има уникално име и се описва с широк спектър от различни характеристики, вкл. речниково тълкование, класификационна информация за неговите синтактични и семантични функции и др.

В рамките на разработката са представени теоретични фреймови модели, съставени от слотове, описващи понятия, процеси и събития от действителността. Словата са описани с един или повече атрибути, които представят знания, описващи съответния елемент. Атрибутите, или така наречените фасети, представят разширени познания за стойностите на слота и позволяват да се контролират неговите стойности [7]. Като атрибути могат да изглеждат и т.нар. „демони“ [8]. Демоните са специален вид процедури (методи), които се стартират автоматично при възникване на определено събитие. Стойността на слот е от определен тип данни. Той може да бъде елементарен или съставен. Съдържанието на всеки слот може да бъде текстова или числова стойност, функция, формула, релация към друг слот, предикатна зависимост или друг фрейм. Структурата на фрейма е динамична, т.е. позволява добавяне на нови слотове във всеки момент [5].

Целта на класифицирането на данните във фреймови структури е да се поддържат и съхраняват данните, да се дефинират връзките помежду им и да се улесни достъпа до тях. Фреймите позволяват класовете да бъдат описани като прототипи на други по-общи класове и тези описания да бъдат организирани в йерархии. Така например студентите могат да се опишат като одушевени хора / представители на класа Човек (Person) плюс набор от свойства, които ги разграничават от други видове учащи лица.

4. Предимства на фреймовия модел

В системите за изкуствен интелект представянето на базата данни под форма на концептуален модел е близо до начина, по който се съхраняват понятията в човешкия мозък. Това предоставя възможност за модулност, гъвкавост и йерархия на данните.

Фреймите за представяне на знания имат следните общи белези: релативност, йерархичност, ключови думи, многофункционалност, асоциативни връзки. Реализирането на идеята за използване на фреймови модели при разработване на информационна система от тип въпрос/отговор ще доведе до няколко положителни ефекта:

- бързо търсене – за всяка част от информацията предварително е отбелязано в кой точно слот на даден фреймов модел се разполага;
- коректно търсене – йерархичното подреждане на фреймите допринася за осъществяване на търсене, което освен стойността на търсения елемент, отчита и аспекта на неговата употреба;
- детайлно търсене - информацията, която получаваме от методите, процедурите и ограниченията за стойностите на атрибутите в слотовете, би позволила разработване на по-сложни алгоритми за представяне и търсене;
- контекстно търсене – различни резултати от търсенето, според различните фреймови модели, свързани към един и същ източник.

Осигуряването на споделени и общи структури на предметната област става съществено и ключово ядро в разработката, а именно: фреймът ще е основен обект в процеса на обмен на информация, и ще се използва за описание на структурата и семантиката на данните.

Представянето на знания в такава компютърно удобна форма ще спомогне за естествено-езиковата обработка и естествено-езиковата генерация на текст.

Езикът за представяне на знания се определя от два аспекта:

- Синтактичен – синтаксисът на един език определя кои конфигурации на компонентите на езика построяват валидни изречения;
- Семантичен – семантиката дефинира кои знания за света се отнасят за изреченията, следователно и твърденията за света, които всяко изречение прави.

Компонентите, на които можем да разделим теоретичния модел за целите на анализа са:

- Лексикална част - определя кои символи или думи се използват в речника на понятията.
- Структурна или синтактична част - описва ограниченията за това как лексемите могат да бъдат подредени в изречения, т.е. граматиката.
- Семантична част - установява начин за асоцииране на реалните значения на света с езиковите компоненти.
- Процедурна част - определя процедурите, които определят как да се създават и модифицират атрибутите, слотовете и фреймите и как да се отговаря на въпроси, използвайки ги.

Въпреки, че основната мотивация за въвеждането на фреймите е семантичното насочване на разсъжденията при анализиращите системи, голяма част от последващата работа върху фрейм-базираните системи е по-скоро върху въпросите за структурното представяне, отколкото върху разсъжденията и логиката. Информацията, съхранявана в фреймите се разглежда като база данни на системата от знания, а контрола на разсъжденията се оставя на back-end частта на системата.

5. Описание на фреймовия модел

В рамките на разработката са представени фреймови модели, съставени от слотове, описващи понятия и ситуации от действителността. Всеки слот е описан с един или повече атрибути, които представят знания, описващи съответния елемент. Атрибутите или така наречените фасети служат за описание както на статично знание (стойност), така и на динамично знание, чрез тъй наречените „прикачени процедури“ (методи) или „демони“ [5]. Демоните са специален вид процедури, които се стартират автоматично при възникване на определено събитие. Стойността на слота е от определен тип данни (елементарен или съставен) [3]. В случаи, когато със слота е асоциирана процедура, той съдържа и инструкции за нейното стартиране.

От информационна гледна точка фреймът е йерархична структура от данни с няколко равнища: име на фрейма на най-високо ниво; полета с описание на обекта или група обекти (слотове); процедури или инструкции.

Групи от семантично свързани фрейми се групират в йерархично подредена система [3], която се състои от фрейми-наследници (част от по-общ фрейм), фрейми и фрейми-родители (йерархично подредени елементи, които изграждат фреймови системи). От тук следва, че фреймите могат да са взаимно свързани помежду си, и да образуват сложни фреймови структури. Счита се, че слотове в сложно организирани фреймови структури могат да станат подфрейми или

фрейми и да формират свои собствени подструктури. Конкретизирането на информацията довежда до създаване на екземпляр във фрейма.

Въз основа на йерархията във фрейма, слотовете могат да бъдат разделени на задължителни(основни) и незадължителни(периферни). Задължителните предават типична конвенционална информация, а незадължителните - специфични характеристики, прилагани в конкретна ситуация или изясняване на информация във връзка с конкретен факт от реалността.

6. Описание на лексикална единица

Лексикална единица е наредената тройка <лема, лексикално значение, примери> [4].

В табл.1 е представен фрейм-прототип „Понятие“ с описание на слотовете (дефинирани с термини на компютърната лингвистика) и тяхното съдържание.

Фреймовият модел в таблица 1. описва отделна лексикална единица, т. нар. „фрейм-прототип“, с подробно описание на неговите слотове и тяхното съдържание. По този начин (фрейми и техни екземпляри) могат да се опишат понятията, използвани в ограничената ПО.

В теоретичния фреймов модел като елементи на разглежданата фреймова система участват различни понятия (concepts) с техните връзки, описани чрез факти и правила от съответната ПО. В структурно отношение той представлява системно обвързани понятия с различно равнище на абстрактност и е първообраз на бъдещата база данни.

Таблица 1. Фрейм-прототип – Понятие – Concept

Име на слот	Обяснение
Name	име на фрейма
Domain	предметна област
Synonims	синонимен ред на името на фрейма
Question Word	въпроси, можещи да се зададат върху името на фрейма
Description	речниково значение
Part Of Speech	синтактична категория
Semantic class	семантичен клас
Sub Frame Of	наследено е от
Has Sub Frames	наследява се от
Core	ядро / корен на думата
BaseForm	основна форма на понятието
Word Options	варианти на думата / словоформи
Gen Answer Word	текст за генериране на отговори
Is Used By	с кои други фрейми е в отношения
Is Argument In	участва като аргумент в предикатна структура

Абстрактният семантичен фрейм според [4] представлява концептуална структура, кодираща културно знание за дадена ситуация и нейните участници. Той представлява съвкупност от елементи, които съдържат необходимата информация, представяща типичното поведение на участниците в ситуацията.

Ще представим базата данни като съвкупност от фреймови модели. От фрейма-прототип, чрез запълване на слотовете с конкретни стойности се получава фрейм-екземпляр.

Таблица 2. Фрейм-прототип Person

Person:		
Име	(текстова стойност - име презиме фамилия)	Кой?
ЕГН/ЛНЧ	(числова стойност)	Какво?
Град	(избор от списък)	Къде?
Адрес	(текстова стойност)	Къде?
Пощ.код	(числова стойност)	Какъв?
Пол	(избор от списък)	Какъв?
Националност	(избор от списък)	Каква?
Sub Frame Of (Person)		
Is Argument In (живее на адрес {човек, адрес}, живее в град{човек, град}, идентичност {студент, ЕГН}, пола му е {човек, пол}, националността е {човек, националност})		

Понятието Student се описва чрез негови основни характеристики.

Таблица 3. Фрейм-прототип Student

Student:		
Име	(Person.Name)	Кой?
Факултетен номер	(текстова стойност)	Какъв?
Семестър	(числова стойност)	Кой?
Успех	(функция)	Колко?
Форма на обучение	(редовно / задочно)	Каква?
Образователна степен	(бакалавър / магистър)	Каква?
Университет	(текстова стойност)	Кой? В кой?
Специалност	(избор от списък)	Каква?
Диплом номер	(текстова стойност)	Каква?
Персонални умения	(текстова стойност)	Какви?
Роля	(избор от списък)	Каква?

Практики	(текстова стойност)	Какви? Каква?
Стажове	(текстова стойност)	Какви? Какъв?
Sub Frame Of	(Person)	
Is Argument In (учи {студент, специалност}, изучава {студент, дисциплина}, е в {студент, курс}, живее в {студент, град}, идентичност {студент, ЕГН})		

Всички стойности от фрейм Person (Person.Name, Person.Address, Person.Nationality и т.н.) се наследяват от фрейм Student, защото Student е вид Person. Student е Person и всички атрибути на Person са атрибути и на Student.

При задаване на въпрос от потребител, ако във въпроса има зададени повече критерии, т.е. ако търсенето е по повече параметри, е удобно същите да се подредят в редица-масив от параметри (параметрите тук са имена на слотове).

7. Описание на фреймови структури

Съществуват два начина на описание на фреймови структури:

- Описание на фреймове, към които принадлежи повече от една лексикална единица / ситуация / процес / събитие / отношение;
- Ориентирани към описание на конкретна лексикална единица и нейните обкръжения;

Таблица 4. Фрейм прототип на процес „Подаване на документ“

	Наименование на слот	Сем. роля / релация	Селективни ограничения	Част на речта	Въпросителна думи
основни елементи	Подател - лице, подаващо документа	Агент	Одушевен, лице, собствено	Съществително	Кой? Чий? На кой?
	Вид на документ	Обект	Неодушевен, нарицателно	Допълнително	Какъв?
	Номер на документ	Темпоратив	Неодушевен, нарицателно	Допълнително	Кой номер? Какъв номер?
	Дата на документ	Темпоратив	Неодушевен, нарицателно	Обстоят. пояснение	Кога?
	Цел / Програма / Кампания	Каузатив	Неодушевен, нарицателно	Съществително	На коя? Каква? С каква цел?
	Получател - лице, приемащо докум.	Рецепиент	Одушевен, лице, собствено	Съществително	Кой? Чий? На кой?

периферни	Отдел -място, където се подава	Локатив	Неодушевен, нарицателно	Обстоят. пояснение	Къде?
	Начин на подаване	Ман	Неодушевен, нарицателно	Обстоят. пояснение	Как?
	Резултат	Резултатив	Неодушевен, нарицателно	Допълнение	Какъв? Какво?
	Статус на документ	Резултатив	Неодушевен, нарицателно	Обстоят. пояснение	Какъв? Какво е състоянието?

Примерът от Таблица 4. отразява логическата функция и мястото на всички елементи от съждението (ситуацията). Използваните критерии и характеристики, които служат за описание на отделните слотове на фрейма, в случая носят информация за:

- Семантичната роля / релация (често срещана и като тематична/ тета роля). Целта на семантичните роли е да обобщят сходни семантични свойства на класове от аргументи. Семантични (тематични, тета) роли са разгледани в [6], където се разграничават следните основни отношения: *агенси, актори, експертиенти, елементативи, обекти, пациенти, адресати, теми, локативи, темпоративи, резултативи и др.*;
- Селективни ограничения – тази класификация разделя съществителните на абстрактни и конкретни, одушевени и неодушевени, лица и не-лица, нарицателни и собствени;
- Част на речта / Синтактична функция – подлог, сказуемо, пряко допълнение, обстоятелствено пояснение;
- Част от синтактична категория - използват се латински букви за означаване на частите на речта, към които принадлежат крайните елементи на изречението (словоформите): Noun – съществително име, Adj – прилагателно име, Verb - глагол, P – предлог, Adv – наречие, Conj – съюз, Pron – местоимение, Num – числително. С S се бележи категорията - изречение.

В следващата табл. 5. са представени фрейми-екземпляри на процеси от тип „Подаване на документи“ с различен характер (кандидатстване за: получаване на здравно осигуряване; настаняване в общежитие; получаване на стипендия; издаване на студентска карта за пътуване с БДЖ; получаване на дубликат на студентска книжка и т.н.).

Таблица 5. Фрейми - екземпляри на процеси от тип „Подаване на документи“

ИД	Подател	Вид	№	Дата	Цел	Служител	Отдел	Начин	Резултат	Статус
1	Костадин Милев	Декларация	241	14.11.2019	Здравно осигуряване	Николина Тодорова	Учебен	На място	Изчаква-ва класиране	Обработка се
2	Камелия Накова	Заявление	1600	07.09.2018	Общежитие	Кирил Станоев	Соц. дейности	На място	Изчаква класиране	Обработка се
3	Стоян Милушев	Заявление	1308	13.01.2019	Стипендия	Кирил Станев	Соц. дейности	Онлайн	Одобрен	Обработен
4	Марин Маринов	Уверение	229	28.03.2019	Да послужи пред БДЖ	Калина Михайлова	Учебен	На място	Издаден	Издаден
5	Мила Ланска	Дубликат	1301	11.02.2018	Дубликат студ. книжка	Мария Накова	Учебен	На място	Издаден	Издаден

Семантичният фрейм „Подаване на документи“ се асоциира с много широк кръг от процеси (подаване на документи за кандидатстване в Университет, за общежитие, за стипендия, за здравни осигуровки, по различни проекти - Еразъм и т. н.), описващи ситуации, при които могат да се обособят различни участници, видове документи, начини за провеждане и резултати. С описанието на фреймовия модел се показва базовата, общата семантична структура, общата ситуация от действителността, към която реферират множество работни процеси.

8. Заключение

Предложената методика се използва за изграждане на информационна система от тип „въпрос-отговор“, с цел извличане на знания в полза на потребители (в случая студенти) и за решаване на комуникационни проблеми. В такъв вид система потребителят задава въпрос на естествен език, а системата може да върне един отговор, списък от отговори или да не успее (ако не намери данни, отговарящи на запитването). Ако въпроса е зададен двусмислено, системата ще поиска доуточняващи въпроси, за да има ефективен, положителен резултат. Източници, от които се извличат данните, е база от йерархично подредени фрейми. Като част от разработката е представен модел за класифициране на понятия, релации и бизнес процеси в конкретна ПО. Моделът ще се анализира и оценява експериментално в следващи изследвания. Разработваните акумулативни фреймови модели за работа със

системи с естествено-езиков интерфейс са замислени като средство, което успешно може да се внедри и в учебна информационна система.

Литература

1. Жекова М., Тотков Г., Панайотова Т. Акумулативни фреймови модели за генериране на тестови единици в системи за е-обучение, Научни трудове на Съюза на учените в България – Пловдив, Серия В. Техника и технологии, том X, 2017.
2. Minsky M., A Framework for Representing Knowledge, MIT, Cambridge, 1974.
3. Шивачева Г., Тотков Г. Акумулативни фреймови модели в обучението по програмиране, октомври 2017.
4. Коева Св., Българският ФреймНет, Семантико-синтактичен речник на българския език, печатно издание, София 2008.
5. Донева Р., Гафтанджиева С, Тотков Г., Фреймови модели в обучението по програмиране, XI Национална конференция „Образованието и изследванията в информационното общество“, 2018.
6. Пенчев Й., Български синтаксис. Управление и свързване, Пловдив 1993;
7. Гафтанджиева С., Донева Р., Тотков Р., Фреймови представяния и приложения, Съюз на учените в България-Пловдив, 2018.
8. Донева Р., Гафтанджиева С, Тотков Г., Фреймови модели в обучението по програмиране, XI Национална конференция „Образованието и изследванията в информационното общество“, 2018.
9. Stefik, M., Bobrow. D.G., Mittal. S.. and Conway, L. Knowledge programming in LOOPS: Report on an experimental course.
10. Aikins, J.S. A representation scheme using both frames and rules. In Rule-Based Expert Systems. B.G. Buchanan and E.H. Shortliffe. Eds. Addison-Wesley, Reading, Mass., 1984.

FRAME MODEL FOR PRESENTATION OF SEATIC ROLLS AND PROCESSES FOR THE ESTABLISHMENT OF NATURAL-LANGUAGE INTERFACE

Maria Jekova, Georgi Totkov

Abstract: This paper presents a research in the field of computer linguistics on a theoretical model for the semantic description of lexical units and processes. The model itself represents knowledge and reflects the conceptual basis of the organization of the memory of a person, as well as its flexibility and visibility. When creating a suitable base of accumulating frame models (AFM), it is necessary to build a linguistically correct and substantiated database structure by describing the concepts, facts, rules, objects and relationships between them, allowing for both flexibility and easier building of the databases, as well as efficient retrieval of different types of information when summarizing the results. In a computer model presentation, the task is to synthesize texts and answer questions in a restricted natural language. The publication is part of a study devoted to the application of frame models for the presentation of knowledge and processes in e-learning.