

**Розробка спеціалізованої бази знань для рекурсивного аналізатора слабо зв'язаної природно-мовної текстової інформації Web- застосувань.**

**Разработка специализированной базы знаний для рекурсивного анализатора слабо связанной естественно-языковой текстовой информации Web-приложений.**

**Developing specialized knowledge bases for recursive parser of weakly connected originated natural language text information and Web-applications.**

**1. Номер державної реєстрації теми - 0110U002409**

**2. Науковий керівник- д.т.н., проф. Стенін О.А, Стенин А.А., Stenin Oleksandr A.**

**3. Суть розробки, основні результати.**

**(укр.)**

База знань з обробки і структуризації природо-мовної інформації, яку здійснює рекурсивний синтаксичний аналізатор, будується на основі нової формальної моделі структурного рівня організації мовного матеріалу, що передбачає рекурсивну схему організації, де чітко визначений елемент рекурсії - базова семантико-синтаксична структура, що описується розширеною моделлю.

Для більшої повноти цього опису використовуються набори метаданих, що створюють багатомірні метаіндекси, які описують не тільки самі мовні об'єкти, а й змістовні фрагменти текстових структурних одиниць, що і створюють додаткові класифікаційні індекси метаданих, які й формують фактично базу знань. Це дозволить проводити паралельну обробку фрагментів тексту рекурсивним синтаксичним аналізатором, підвищити швидкість обробки текстів природо-мовної інформації та уніфікувати процеси їх обробки.

Проектування класифікаторів бази знань і розробка відповідного пакету запитів для пошуку всіх об'єктів (текстів) або їх фрагментів, що відносяться до поняття або ситуації (рубриці) класифікатора, забезпечує можливість повної автоматизації збору і рубрикації інформації із внутрішніх джерел.

Таким чином, пропонується вирішення проблеми автоматизації формування обмеженої кількості структурних моделей за рахунок значного скорочення структур представлення природно-мовної інформації, а також за рахунок заміщення структурного аналізу – аналізом параметричних моделей на основі метаданих, що значно спрощує синтаксичний аналіз і надає можливість оперативної підтримки в реальному часі систематизованих баз даних і знань та подальшого мовного синтезу.

**(рос.)**

База знаний по обработке и структуризации естественно - языковой информации, которую осуществляет рекурсивный синтаксический анализатор, строится на основе новой формальной модели структурного уровня организации языкового материала, который предусматривает рекурсивную схему организации, где четко определен элемент рекурсии - базовая семантико-синтаксична структура, которая описывается расширенной моделью.

Для большей полноты этого описания используются наборы метаданных, которые создают многомерные метаиндексы, которые описывают не только сами языковые объекты, но и содержательные фрагменты текстовых структурных единиц, которые и создают дополнительные классификационные индексы метаданных, и на их основе формируют фактически базу знаний. Это позволяет проводить параллельную обработку фрагментов текста рекурсивным синтаксическим анализатором, повысить скорость обработки текстов естественно - языковой информации и унифицировать процессы их обработки.

Проектирование классификаторов базы знаний и разработка соответствующего пакета запросов для поиска всех объектов (текстов) или их фрагментов, которые относятся к понятию или ситуации (рубрике) классификатора, обеспечивают возможность полной автоматизации сбора и рубрикации информации из внутренних источников.

Таким образом, предлагается решение проблемы автоматизации формирования ограниченного количества структурных моделей за счет значительного сокращения структур

представления естественно- языковой информации, а также за счет замещения структурного анализа - анализом параметрических моделей на основе метаданных. Это значительно упрощает синтаксический анализ и предоставляет возможность оперативной поддержки в реальном времени систематизированных баз данных и знаний, последующего языкового синтеза.

(англ.)

Knowledge base for processing and structuring of natural - language information, which provides a recursive text parser, based on a new formal model of the structural level of organization of linguistic material, which provides a recursive scheme of organization that clearly defined an element of recursion - the basic semantic and syntactic structure, which is described by the extended model.

For a fuller this description used the metadata sets, that create a multidimensional metaindexes that describe not only the linguistic objects and fragments of text content of structural units, but which create additional metadata classification numbers, and on that basis in fact form the knowledge base. This allows parallel processing of text parts of the recursive text parser, increase the speed of text processing natural - language information and unify their processing.

Designing classifier knowledge base and develop an appropriate package of requests to search for all objects (text) or their fragments provide the ability to fully automate the collection and categorization of information from internal sources.

Offer a solution to the problem of automating the formation of a limited number of structural models due to a significant reduction in the structures of natural-language information representation, as well as by substitution of structural analysis - the analysis of parametric metadata models.

**4. Наявність охоронних документів на об'єкти права інтелектуальної власності. – не має**  
**5. Порівняння зі світовими аналогами.**

Робота відповідає світовому рівню з напрямку Business Decision-Makers, який обумовлює склад реальних компонентів ІТ-систем, в тому числі накопичення та узагальнення інформації, її розпізнавання та класифікацію, вимоги до функціонування таких систем, отримання нових знань з технології обробки природно-мовної текстової інформації за схемою рекурсивної організації її структури на основі обмеженої кількості структурних шаблонів, а також моделей для їх реалізації, що надасть можливість значно спростити та прискорити синтаксичний аналіз текстів.

**6. Економічна привабливість для просування на ринок.**

Застосування розроблених моделей та технології за рахунок заміщення структурного аналізу – аналізом параметричних моделей на основі метаданих, значно спрощує синтаксичний аналіз і надає можливість оперативної підтримки в реальному часі систематизованих баз даних і знань та подальшого мовного синтезу, дозволяє проводити паралельну обробку фрагментів тексту рекурсивним аналізатором, підвищити швидкість обробки текстів природно-мовної інформації та уніфікувати процеси їх обробки.

Ефективність розробки заключається у зменшенні на порядок трудомісткості аналізу різних текстових інформаційних джерел, у спрощеності пошуку в ІТ-системах загального використання, що призводить до *економії трудових витрат* користувачів *і коштів* для комп'ютерних ресурсів.

**7. Потенційні користувачі (галузі, міністерства, підприємства, організації).**

Розроблені моделі та засоби можуть знайти застосування у таких організаціях:

- Інститут програмних систем та інститут проблем з інформації НАН України;
- В ІТ-галузі на підприємствах, які працюють в області аналізу і моделювання мовного поведіння людини та їх дослідження з метою автоматизації вводу інформації;
- Структурні підрозділи НТУУ «КПІ», в тому числі УПТО, КБ ІС, ФАКС, ФІОТ, ФПМ, при створенні наукоємного інтелектуального комплексу для дистанційного навчання і підготовки бакалаврів і магістрів університету по вбудованим системам (embedded systems);

- Вітчизняними та закордонними компаніями при розробці інтелектуальних систем голосового управління при взаємодії людини і комп'ютера.

## **8. Стан готовності розробки.**

Розроблено комп'ютерні моделі та відпрацьована технологія рекурсивної обробки поліпредикатних структур тексту, а також архітектура бази знань з їх лінгвістичного аналізу. Можлива розробка пілотного дослідного зразка такої БЗ з відповідними компонентами обробки технічних текстів.

## **9. Існуючі результати впровадження.**

Основні положення роботи впроваджені при підготовці фахівців із спеціальності 7.091402 "Гнучкі комп'ютеризовані системи та робототехніка" в розрізі створення нових розділів по курсах "Інформаційні природно-мовні технології" (розділ "Рекурсивна структура організації мовної інформації"), "Корпоративні системи та технології" (розділ "Рекурсивні процеси обробки текстової інформації в пошукових системах") , "Сучасні інформаційні технології" (розділ "Сучасні технології обробки текстової інформації"). Матеріали роботи використані при підготовці докторської дисертації доц. Кисленко Ю.І. "Створення інформаційного ядра природньо-мовних технологій".

Заплановано сумісне використання моделей та технології рекурсивного аналізу текстів при створенні системи віртуальних архівів для корпоративної мережі ВНЗ в рамках Державної програми "Інформаційні та комунікаційні технології в освіті і науці" для системи «Електронний кампус НТУУ «КПІ», а також компаніями «Фонд качественной политики» (Україна) та «REAL TIME INTELLIGENCE» (США).

## **10. Назва організації, телефон, E-mail**

НДІ системних технологій НТУУ «КПІ», Факультет інформатики та обчислювальної техніки, кафедра технічної кібернетики, (044) 406-86-13, e-mail: [y.timoshin@kpi.ua](mailto:y.timoshin@kpi.ua)

## **11. Перелік публікацій за матеріалами досліджень за період виконання розробки**

1. Кисленко Ю.І. Мовленнєва діяльність та її моделювання // Адаптивні системи автоматичного управління - Дніпропетровськ: ДНВП Системні технології, 2009-Вип.16(36).-С.49-57
2. Тимофеев В.І., Тимошин Ю.А., Ярченко В.П., Шуст С.О. Архітектура систем обробки та архівації бібліотечного контенту в гетерогенному середовищі ВНЗ // Інформаційні технології в освіті: збірник наук. праць, вип.4.-Херсон: вид. ХДУ, 2009.- с. 41-48
3. Кисленко Ю.І. Інформаційний підхід до аналізу структурного рівня мовної організації // Сб. трудов междунар. конф. «Языковые технологии в современном мире», Ялта, 2010, стр. 90-101
4. Стенін О.А., Тимошин Ю.А., Шемсединов Т.Г., Шуст С.О. Розробка фізичних і логічних метрик в задачі багатокритеріальної оптимізації інформаційного навантаження при структуризації корпоративного центру даних // Адаптивні системи автоматичного управління - Дніпропетровськ: ДНВП Системні технології, 2009-Вип.15(35).-С.97-102
5. Стенін О.А., Тимошин Ю.А., Ярченко В.П., Шуст С.О. Вирішення задачі багатокритеріальної оптимізації інформаційного навантаження при структуризації корпоративного центру даних // Адаптивні системи автоматичного управління - Дніпропетровськ: ДНВП Системні технології, 2010-Вип.16(36).-с.135-141
6. Шемсединов Т.Г.. Метамоделі в задачах интеграции информационных систем // Адаптивні системи автоматичного управління - Дніпропетровськ: ДНВП Системні технології, 2010- Вип.17(37).-С.99-103
7. Кисленко Ю.І., Черевко О.С. Категорії часу та простору в інформаційних природно-мовних технологіях. // Адаптивні системи автоматичного управління - Дніпропетровськ: ДНВП Системні технології, 2011- Вип.18(38).-С.62-70
8. Кисленко Ю.І., Терентьев А.В. Проблемы и перспективы развития поисковых систем // Сборник трудов конференции «Искусственный интеллект – 2011», т.3, Казивели , сентябрь 2011. – С.55-66.

9. Стенін О.А., Тимошин Ю.А., Галаган В.Г., Коломієц Д. Про функціональність бази знань з аналізу та обробки лінгвістичної інформації // Адаптивні системи автоматичного управління - Дніпропетровськ: ДНВП Системні технології, 2012-Вип.20(40).-с.
10. Стенін О.А., Тимошин Ю.А., Кисленко Ю.І., Шемседінов Т.Г. Архітектура бази знань для лінгвістичного аналізу природо-мовних текстів // Адаптивні системи автоматичного управління - Дніпропетровськ: ДНВП Системні технології, 2012-Вип.20(40).-с.