

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/305787278>

# About NAddressing (К вопросу о естественно-языковой адрессации)

Book · September 2012

CITATIONS

9

READS

59

## 3 authors:



[Krassimira Borislavova Ivanova](#)

University of Telecommunications and Post

42 PUBLICATIONS 52 CITATIONS

SEE PROFILE



[Vitalii Velychko](#)

52 PUBLICATIONS 123 CITATIONS

SEE PROFILE



[Krassimir Markov](#)

Institute of Information Theories and Applications

58 PUBLICATIONS 202 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



Information Quality View project



MDNDB - Multi-Dimensional Numbered Data Base View project

National Academy of Sciences of Ukraine  
V.M.Glushkov Institute of Cybernetics

---

---

I T H E A

---

---



# Problems of Computer Intellectualization

---

---

KYIV - SOFIA

---

---

2 0 1 2



**National Academy of Sciences of Ukraine  
V.M.Glushkov Institute of Cybernetics**

**I T H E A**

**Problems  
of  
Computer  
Intellectualization**

**Kyiv - Sofia**

**2 0 1 2**

**Vitalii Velychko, Alexey Volosyn, Krassimir Markov (ed.)**

**Problems of Computer Intellectualization**

V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine,  
ITHEA®

2012, Kyiv, Ukraine - Sofia, Bulgaria

ISBN: 978-966-02-6529-5 (printed)

ISBN: 978-954-16-0061-0 (printed)

ISBN: 978-954-16-0062-7 (online)

ITHEA IBS ISC No.: 28

First edition

Printed in Ukraine

Recommended for publication by The Scientific Council of the Institute of Information Theories and Applications FOI ITHEA

This issue contains a collection of papers in the fields of Natural Language Processing, Pattern Recognition and Decision Making as well as other important areas of Artificial Intelligence. Papers in this issue are selected from the XVIII-th International Conference "Knowledge-Dialogue-Solution", Kyiv, Ukraine, 2012 – a part of the Joint International Events of Informatics "ITA 2012", Autumn Session.

It is represented that book articles will be interesting for experts in the field of information technologies as well as for practical users.

**© All rights reserved.**

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

**Copyright © 2012**

© 2012 V.M.Glushkov Institute of Cybernetics of NASU – Publisher;

© 2012 ITHEA® – Publisher;

© 2012 Vitalii Velychko, Alexey Volosyn, Krassimir Markov – Editors

© 2012 For all authors in the book.

® ITHEA is a registered trade mark.

**ISBN: 978-966-02-6529-5 (printed)**

**ISBN: 978-954-16-0061-0 (printed)**

**ISBN: 978-954-16-0062-7 (online)**

C\o Jusautor, Sofia, 2012

## PREFACE

ITHEA International Scientific Society (ITHEA ISS) is aimed to support growing collaboration between scientists from all over the world. The scope of the books of the ITHEA ISS covers the area of Informatics including Computer Science. ITHEA ISS welcomes scientific papers and books connected with any information theory or its application.

ITHEA ISS rules for preparing the manuscripts are compulsory.

ITHEA Publishing House is the official publisher of the works of the ITHEA ISS.

Responsibility for papers and books published by ITHEA belongs to authors.

This book maintains articles on actual problems of research and application of information technologies, especially the new approaches, models, algorithms and methods for information modeling of knowledge in: Intelligent NL Processing; Management of computing and information system; Knowledge Engineering; Mathematical Foundation of Artificial Intelligence; Computing; Decision Making; Economics Decision Support Systems; Modern (e-) Learning; Philosophy and Methodology of Informatics.

Papers in this issue are peer reviewed and selected from the XVIII-th International Conference "Knowledge-Dialogue-Solution", Kyiv, Ukraine, 2012 – a part of the ITHEA Joint International Events of Informatics "ITA 2012", Autumn Session.

KDS 2012 has been organized by V.M.Glushkov Institute of Cybernetics of National Academy of Sciences of Ukraine in collaboration with:

- ITHEA International Scientific Society
- Association of Developers and Users of Intelligent Systems (Ukraine)
- Federation of the Scientific - Engineering Unions /FNTS/ (Bulgaria)
- National Academy of Sciences of Ukraine
- Institute of Information Theories and Applications FOI ITHEA (Bulgaria)
- Taras Shevchenko National University of Kiev (Ukraine)
- Kharkiv National University of Radio Electronics (Ukraine)
- Institute of Mathematics and Informatics, BAS (Bulgaria)
- Institute of Mathematics of SD RAN (Russia)

The International Conference "Knowledge-Dialogue-Solution" (KDS) was established as result of the scientific co-operation organized within 1986-1992 by two international workgroups (IWG) researching the problems of data bases and artificial intelligence. As a result of tight relation between these problems in 1990 in Budapest appeared the International scientific Working Group of "Data Base Intellectualization" (IWGDBI) integrating the possibilities of databases with the creative process support tools. Heads of the IWGDBI were V.Gladun (Ukraine) and R.Kirkova (Bulgaria).

In 1992 the "Association of Developers and Users of Intelligent Systems" (ADUIS) had been established. ADUIS was chosen to be the main organization of the KDS international conferences. In the same time the "Institute of information theories and application FOI ITHEA" (ITHEA) started working in Bulgaria.

During the years, in collaboration with ITHEA, KDS had been organized in Ukraine, Poland, Russia, and Bulgaria.

More information about ITHEA International Conferences is given at the [www.ithea.org](http://www.ithea.org).

We express our thanks to all authors, editors and collaborators.

The great success of ITHEA International Journals, Books and Conferences belongs to the whole of the ITHEA International Scientific Society.

*Kyiv-Sofia, September 2012*

*V. Velychko, A. Voloshin, Kr. Markov*

## TABLE OF CONTENTS

<i>Preface</i> .....	3
<i>Table of Contents</i> .....	5
<i>Index of Authors</i> .....	8
<b>Intelligent NL Processing</b>	
Формальное определение ситуации для семантического анализа текстов естественного языка <i>Леонид Святогор, Виктор Гладун</i> .....	9
О моделировании понимания <i>Владимир Сторож</i> .....	21
Численные меры “сплоченности” именных групп <i>Леонид Леоненко</i> .....	31
<b>Management of computing and information system</b>	
Method for Determination of Interrelation between Access Network Characteristics <i>Galyna Gayvoronska, Svetlana Sakharova</i> .....	41
About the Problem of Development of Messages Stream Model in the Convergent Telecommunication Network <i>Galyna Gayvoronska, Maxim Solomitsky</i> .....	47
Modeling Telecommunications Traffic Using the Stochastic Multifractal Cascade Process <i>Lyudmyla Kirichenko, Tamara Radivilova, Eias Kayali</i> .....	55
Steps in the Development of the Information Networks’ User Model as Badly Formalized Object <i>Anastasia Smirnova</i> .....	64
Анализ эффективности алгоритмов резервирования ресурсов оптической сети <i>Юрий Гриньков</i> .....	71
<b>Knowledge Engineering</b>	
К вопросу естественно-языковой адресации <i>Крассимира Иванова, Виталий Величко, Крассимир Марков</i> .....	77
Интеллектуализация экспертных систем с помощью онтологий <i>Глибовец Н.Н., Красиков Д.С.</i> .....	84
К вопросу разработки информационно-справочного обеспечения инструментального комплекса онтологического назначения <i>Палагин А.В., Петренко Н.Г., Величко В.Ю., Тихонов Ю.Л., Семенов В.В., Митрофанова А.Е.</i> .....	91
Онтологоуправляемая системная оптимизация <i>Юрий Чаплинский, Елена Субботина</i> .....	99
<b>Mathematical Foundation of Artificial Intelligence</b>	
‘Feature Vectors’ in Grouping Information Problem in Applied Mathematics: Vectors and Matrixes <i>Donchenko V., Zinko T., Skotarenko F.</i> .....	111



Алгоритм построения выпуклого продолжения полиномов на полиперестановках и сфера его применения <i>Пичугина Оксана</i> .....	125
---	-----

## Computing

System of Programs Proving <i>Alexander Letichevsky, Olexander Letichevskiy, Marina Morokhovets, Vladimir Peschanenko</i> .....	133
Обучение рекуррентных нейронных сетей методом псевдорегуляризации для многошагового прогнозирования на примере хаотического процесса Маккея-Гласса <i>Артем Чернодуб</i> .....	141
Анализ алгоритмов прогнозирования третичной структуры протеина на базе метода оптимизации муравьиными колониями <i>Леонид Гуляницкий, Виталина Рудык</i> .....	152
Combining the Backpropagation Algorithm Training Functions for a Bitmap-Character Recognition Problem under Noise with Feed-Forward Neural Network <i>Vadim Romanuke</i> .....	160
Создание специализированной поисковой системы на базе облачных технологий <i>Глибовец Андрей, Ситмамбетов Назим</i> .....	167
Создание рекомендационной системы учебного типа с использованием фреймворка Windows Communication Foundation <i>Глибовец Николай Николаевич, Сидоренко Марина Олеговна</i> .....	176
Аспекты неклассической теории номинации и их использование в формальных языках <i>Россада Татьяна</i> .....	182

## Decision Making

Многокритериальная оценка альтернатив <i>Альберт Воронин</i> .....	190
Expert System of Rational Decision Making of Distribution of Electric in Branched Electrical Network <i>Kyzemin Oleksandr, Irina Gurina</i> .....	200

## Economics Decision Support Systems

Анализ кредитоспособности заемщиков - юридических лиц в условиях неопределенности <i>Юрий Зайченко</i> .....	208
Обобщения принципов распределения затрат для нечетких моделей распределения <i>Алексей Волошин, Василий Лавер</i> .....	219
Разработка и использование интегрального индекса бюджетной сферы <i>Леонид Гуляницкий, Диана Омелянчик</i> .....	227
Двухуровневая модель нечеткого рационального многокритериального выбора <i>Н.Н.Маляр, В.В. Полищук</i> .....	242
Оценка эффективности прогнозирования и принятия решений на финансовом рынке <i>Александр Берзлев</i> .....	249

Модель оптимизации прибыли предприятий лесопильной отрасли <i>Роман Шулла, Михаил Повидайчик</i> .....	258
<b>Modern (e-) Learning</b>	
Формирование базовых структур восприятия информации <i>Владимир Донченко</i> .....	263
Почему не используются электронные образовательные ресурсы <i>Андрей Федосеев</i> .....	269
Образовательные парадигмы в цифровую эпоху <i>Диана Богданова</i> .....	274
Электронное портфолио как средство сопровождения студентов в образовательном процессе вузов <i>Воронцов Сергей, Болгов Михаил, Артемьева Ирина</i> .....	279
Учеба в виртуальных мирах <i>Григоркив Василий, Винничук Елена, Винничук Игорь</i> .....	287
Учебно-методическая тестирующе-оценивающая программная система поддержки учебных курсов по теории принятия решений <i>Алексей Волошин, Даниил Ковалёв</i> .....	293
Использование организационной модели ОрегА для управления рынком знаний <i>Глибовец Николай Николаевич, Сидоренко Марина Олеговна</i> .....	299
Процессы выбора альтернативных решений при проектировании информационных сетей <i>Галина Гайворонская</i> .....	307
Опыт использования методов решения задач выбора в условиях неопределенности и риска при проектировании компьютерных сетей <i>Максим Соломицкий, Оксана Болотина</i> .....	318
Применение обучающей программы SMPR при выборе технологий сетей доступа <i>Антон Бондаренко</i> .....	332
Моделирование оценки знаний выпускника вуза <i>Николай Маляр, Анатолий Штимак</i> .....	339
Некоторые аспекты использования информационных технологий в образовательной системе Украины <i>Наталья Кондрук, Наталья Юрченко</i> .....	344
<b>Philosophy and Methodology of Informatics</b>	
Когнитивная семиотика в процессах образного мышления <i>Юрий Валькман</i> .....	349
Divergent and Multiple-Valued Sequences and Functions <i>Igor Gorban</i> .....	358
ITHEA International Scientific Society .....	374

## INDEX OF AUTHORS

Volodymyr	Donchenko	111	Юрий	Гриньков	71
Galyna	Gayvoronska	41, 47	Леонид	Гуляницкий	152, 227
Igor	Gorban	358	Владимир	Донченко	263
Irina	Gurina	200	Юрий	Зайченко	208
Eias	Kayali	55	Крассимира	Иванова	77
Lyudmyla	Kirichenko	55	Даниил	Ковалёв	293
Oleksandr	Kyzemin	200	Наталия	Кондрук	344
Olexander	Letichevskiy	133	Дмитрий	Красиков	84
Alexander	Letichevsky	133	Василий	Лавер	219
Marina	Morokhovets	133	Леонид	Леоненко	31
Vladimir	Peschanenko	133	Николай	Маляр	242, 339
Tamara	Radivilova	55	Крассимир	Марков	77
Vadim	Romanuke	160	Анна	Митрофанова	91
Svetlana	Sakharova	41	Диана	Омельянчик	227
Fedir	Skotarenko	111	Александр	Палагин	91
Anastasia	Smirnova	64	Николай	Петренко	91
Maxim	Solomitsky	47	Оксана	Пичугина	125
Taras	Zinko	111	Михаил	Повидайчик	258
Ирина	Артемьева	279	Владимир	Полищук	242
Александр	Берзлев	249	Татьяна	Россада	182
Диана	Богданова	274	Виталина	Рудык	152
Михаил	Болгов	279	Леонид	Святогор	9
Оксана	Болотина	318	Виталий	Семенов	91
Антон	Бондаренко	332	Марина	Сидоренко	176, 299
Юрий	Валькман	349	Назим	Ситмамбетов	167
Виталий	Величко	77, 91	Максим	Соломицкий	318
Елена	Винничук	287	Владимир	Сторож	21
Игорь	Винничук	287	Елена	Субботина	99
Алексей	Волошин	219, 293	Юрий	Тихонов	91
Альберт	Воронин	190	Андрей	Федосеев	269
Сергей	Воронцов	279	Юрий	Чаплинский	99
Галина	Гайворонская	307	Артем	Чернодуб	141
Виктор	Гладун	9	Анатолий	Штимак	339
Андрей	Глибовец	167	Роман	Шулла	258
Николай	Глибовец	84, 176, 299	Наталия	Юрченко	344
Василий	Григоркив	287			

---

---

## Knowledge Engineering

---

---

### К ВОПРОСУ ЕСТЕСТВЕННО-ЯЗЫКОВОЙ АДРЕСАЦИИ

**Крассимира Иванова, Виталий Величко, Крассимир Марков**

**Аннотация:** В настоящей работе представлена идея естественно-языковой адресации. Это дополнительная возможность для представления онтологической информации в интеллектуальных системах. Естественно-языковая адресация имеет ряд преимуществ. На первом месте – это линейная алгоритмическая сложность, которая зависит от максимальной длины слов ( $max\_L$ ), а не от их количества. Во-вторых, это уменьшение объема занимаемой памяти – дополнительные индексы не используются. В-третьих, уменьшение времени обработки из-за полного отсутствия поиска – информация извлекается прямо по адресу. Необходимо отметить, что это универсальное представление информации одновременно доступной как для человека, так и для автоматизированных систем. Такой способ организации информации применим для ее хранения и использования в библиотеках онтологий, терминов, понятий, текстовых документов.

**Ключевые слова:** Естественно-языковая адресация, организация онтологических баз данных.

**ACM Classification Keywords:** D.4.2 Storage Management; E.2 Data Storage Representations

---

#### Введение

---

Растущее развитие средств для оперирования с онтологиями и базами знаний определяет устойчивую тенденцию увеличения объема мета-информации, создаваемой людьми и компьютерами. С момента появления Semantic Web [sw, 2012] практически любой объект в Интернете содержит значительное количество метаданных, описывающих различные семантические аспекты. Наличие стандартизированных языков для описания метаданных (XML, SHOE, DAML+OIL, RDF и RDFS и др.) делает мета-информацию доступной для анализа и интерпретации.

Увеличение количества метаданных является феноменом, который требует особого внимания и анализа. Бытует иллюзорное впечатление, что в глобальной сети уже есть все, что нас интересует, но пока не описано полностью мета-данными и является вопросом времени, чтобы это произошло. Когда это случится, семантико-поисковые системы будут находить для нас необходимую информацию, которая должна быть и "достаточной".

Более чем очевидно, что "Никто не обнимет необъятного" [Prutkov, 1863], особенно если к нему добавлено "мета-необъятное". Поэтому поиск эффективных решений для работы с естественными структурами языка и их смыслом остается серьезной проблемой. Наше внимание в этой работе направлено на поиск новых возможностей хранения онтологических структур, основанных на специфической "естественно-языковой адресации".

### Пример адресации в системе WordNet

**WordNet** — это семантическая сеть для английского языка, разработанная в Принстонском университете, и свободно доступная вместе с сопутствующим программным обеспечением [WordNet, 2012]. WordNet можно свободно использовать в коммерческих и научных целях. Для работы с ним существует несколько программ, множество интерфейсов и API, реализованное как на большинстве возможных языков, так и с помощью протокола DICT, программы GoldenDict и других. Пакеты WordNet присутствуют в некоторых репозиториях GNU ПО, Linux и их дистрибутивах.

Рассмотрим организацию информации в системе WordNet. Для примера мы выбрали слово „accession”.

На запрос о слове „accession”, система WordNet выдает следующую информацию (рис. 1):

The noun accession has 6 senses (no senses from tagged texts)

1. {13251723} <noun.process> accession#1 -- (a process of increasing by addition (as to a collection or group); "the art collection grew through accession")
2. {13170404} <noun.possession> accession1#2 -- ((civil law) the right to all of that which your property produces whether by growth or improvement)
3. {13082910} <noun.possession> accession#3, addition#4 -- (something added to what you already have; "the librarian shelved the new accessions"; "he was a new addition to the staff")
4. {07078650} <noun.communication> accession2#4, assenting#1 -- (agreeing with or consenting to (often unwillingly); "accession to such demands would set a dangerous precedent"; "assenting to the Congressional determination")
5. {05115154} <noun.attribute> entree#2, access#1, accession#5, admittance#1 -- (the right to enter)
6. {00232781} <noun.act> accession3#6, rise to power#1 -- (the act of attaining or gaining access to a new office or right or position (especially the throne); "Elizabeth's accession in 1558")

The verb accession has 1 sense (no senses from tagged texts)

1. {00989696} <verb.communication> accession#1 -- (make a record of additions to a collection, such as a library)

Рис. 1 Ответ системы WordNet на запрос о слове „accession”

Рассмотрим, как именно получен этот ответ в системе WordNet. Словарь WordNet состоит из 4 сетей для основных знаменательных частей речи: существительных, глаголов, прилагательных и наречий. Базовой словарной единицей в WordNet является не отдельное слово, а так называемый синонимический ряд («синсеты»), объединяющие слова со схожим значением и по сути своей являющимися узлами сети. Для удобства использования словаря человеком каждый синсет дополнен определением и примерами употребления слов в контексте. Слово или словосочетание может появляться более чем в одном синсете и иметь более одной категории части речи. Каждый синсет содержит список синонимов или синонимичных словосочетаний и указатели, описывающие отношения между ним и другими синсетами. Слова, имеющие

несколько значений, включаются в несколько синсетов и могут быть причислены к различным синтаксическим и лексическим классам.

WordNet хранит информацию в четырех основных файлах данных (существительные, глаголы, прилагательные и наречия). В каждом из этих файлов структура данных одинакова – для каждого слова хранится один или несколько наборов синонимических множеств (синсетов), доступ к которым осуществляется по адресу первого байта синсета, который задан явно с помощью восьми десятичных знаков записанных с первого байта синсета (рис. 2 и рис. 3). Данные в синсете разделены пробелами. После адреса синсета следуют три служебных поля, а вслед за ними само слово, после него следует другая лингвистическая информация. Необходимо отметить, что ссылки на другие синсеты задаются снова через их абсолютные адреса в файле. (Элементы, которые нас интересуют, выделены жирным шрифтом).

```

13251723 22 n 01 accession 0 001 @ 13323403 n 0000 | a process of
  increasing by addition (as to a collection or group); "the art
  collection grew through accession"

13170404 21 n 01 accession 1 002 @ 13070995 n 0000 ;c 08338303 n
  0000 | (civil law) the right to all of that which your property
  produces whether by growth or improvement

13082910 21 n 02 accession 0 addition 0 001 @ 13082742 n 0000 |
  something added to what you already have; "the librarian
  shelved the new accessions"; "he was a new addition to the
  staff"

07078650 10 n 02 accession 2 assenting 0 002 @ 07076600 n 0000 +
  00795631 v 0102 | agreeing with or consenting to (often
  unwillingly); "accession to such demands would set a dangerous
  precedent"; "assenting to the Congressional determination"

05115154 07 n 04 entree 0 access 0 accession 0 admittance 0 003 @
  05113619 n 0000 + 02426186 v 0401 ~ 05119817 n 0000 | the right
  to enter

00232781 04 n 02 accession 3 rise_to_power 0 003 @ 00060914 n 0000 +
  01989112 v 0101 + 02358456 v 0101 | the act of attaining or
  gaining access to a new office or right or position (especially
  the throne); "Elizabeth's accession in 1558"

```

Рис. 2. Синсеты слова „accession” в WordNet файле с данными для существительных

```

00989696 32 v 01 accession 0 002 @ 00990286 v 0000 ;c 00897092 n
  0000 01 + 08 00 | make a record of additions to a collection,
  such as a library

```

Рис. 3. Синсеты слова „accession” в WordNet файле данных для глаголов

Для нас важно, как выполняется доступ к данному синсету. Очевидно, что необходимо где-то хранить информацию о том, где находятся синсеты для каждого слова. Это делается через WordNet индексные файлы, которых тоже четыре, в соответствии с файлами данных (существительные, глаголы, прилагательные и наречия). Они отсортированы в алфавитном порядке слов, каждому слову соответствует одна строка в индексном файле, содержащая слово, краткую служебную информацию и абсолютные адреса всех синсетов, в которые входит слово (рис. 4 и рис. 5).

```
accession n 6 4 @ ~ + ; 6 0 13251723 13170404 13082910 07078650
05115154 00232781
```

Рис. 4. Строка для слова „accession” в WordNet индексном файле с данными для существительных

```
accession v 1 2 @ ; 1 0 00989696
```

Рис. 5. Строка для слова „accession” в WordNet индексном файле с данными для глаголов

Чтобы прочитать все синсеты для данного слова, сначала выполняется бинарный поиск во всех индексных файлах, а затем через прямой доступ по абсолютным адресам читаются данные из файлов. Алгоритмическая сложность в данном случае  $O(n_n * \lg(n_n) + n_v * \lg(n_v) + n_a * \lg(n_a) + n_r * \lg(n_r))$ , где  $n_n$ ,  $n_v$ ,  $n_a$  и  $n_r$  представляют собой количества существительных, глаголов, прилагательных и наречий.

Существует и второй способ получения абсолютных адресов синсетов. Он выполняется с помощью т.н. индекса смыслов (sense index). Этот индекс тоже отсортирован, но каждое слово записывается в таком количестве строк, сколько синсетов существуют для этого слова во всех файлах данных. Например, слово „accession” имеет семь строк в индексе смыслов - шесть для его значений как существительное и одна - в качестве глагола. Каждая строка содержит только один абсолютный адрес синсета в соответствующем файле данных (рис. 6).

```
accession%1:04:03:: 00232781 6 0
accession%1:07:00:: 05115154 5 0
accession%1:10:02:: 07078650 4 0
accession%1:21:00:: 13082910 3 0
accession%1:21:01:: 13170404 2 0
accession%1:22:00:: 13251723 1 0
accession%2:32:00:: 00989696 1 0
```

Рис. 6 Строки слова „accession” в индексе смыслов

Для получения всех синсетов данного слова в индексе смыслов, выполняется сначала бинарный поиск, затем просматриваются все имеющиеся для данного слова строки, и, наконец, по всем полученным адресам непосредственно считывают синсеты из файлов данных. Алгоритмическая сложность в данном случае больше, чем  $O(n * \lg(n))$ , где  $n = n_n + n_v + n_a + n_r$ , т.е.  $n$  - общее количество слов в базе данных (существительные+глаголы+прилагательные+наречия), так как из-за множества значений, слова могут повторяться много раз, что приводит к увеличению количества операций по нахождению всех строк этого слова.

### Естественно-языковая адресация

Организация информации в системе WordNet позволяет быстро получить необходимую информацию, используя „одновременный” бинарный поиск в четырех стандартных индексах или в одном индексе смыслов, а затем выполнить позиционирование непосредственно в соответствующих файлах данных.

В данном случае необходимо отметить некоторые недостатки такой организации:

- 1) абсолютная адресация удобна для компьютерной обработки, но не удобна для пользователя;
- 2) ручное формирование абсолютных адресов невозможно, а их использование не может быть реализовано без поддержки соответствующей программы;
- 3) пользователю предоставляется статическая ("скомпилированная") версия базы данных, которую невозможно развивать - она пригодна только для чтения.

Любые изменения, приводящие к изменению количества байтов в основном файле данных, делает его непригодным для использования из-за абсолютных адресов, являющимися указателями. Например, на рис. 7 показан синсет слова „accession” из (а) настоящей и (б) более ранней версии WordNet файлов данных для существительных. Более ранняя версия синсета опубликована в [Palagin et al, 2011]. На рис. 7 (а) и (б) видна разница в адресах.

```
13082910 21 n 02 accession 0 addition 0 001 @ 13082742 n 0000 |
something added to what you already have; "the librarian
shelved the new accessions"; "he was a new addition to the
staff"
```

а)

```
00047131 04 n 02 accession 0 addition 0 001 @ 09536731 n 0000 |
something added to what you have already; "the librarian
shelved the new accessions"; "he was a new addition to the
staff"
```

б)

Рис. 7 Синсет слова „accession” из (а) настоящей и (б) более ранней версии файла для существительных. Это означает, что любое изменение в информации, даже на один байт, требует полной рекомпиляции соответствующей части базы данных – основного файла, его индекса, индекса смыслов, который соответствует всем основным файлам.

Если проанализируем структуру синсета, увидим одну очень важную особенность.

В памяти компьютера все символы представляются цифровыми кодами (занимающими один, два или четыре байта, в зависимости от системы кодирования - ASCII или Unicode). Таким образом, один уникальный цифровой код (абсолютный адрес) указывает на другой, тоже уникальный цифровой код (компьютерное представление слова, например в ASCII-кодировке слово „accession” имеет следующее представление: 97 99 99 101 115 115 105 111 110).

Этого можно избежать, если использовать другой тип организации информации.

Внутреннее представление (код) слова может быть непосредственно использовано при построении онтологии. Этот код можно рассматривать как пространственный адрес (в девятимерном пространстве в случае со словом „accession”). По адресу слова можно сохранить всю информацию синсета и получить ее снова через этот адрес.

Для людей адрес слова „accession” будет представлен самим словом „accession”, а для компьютера через вектор (97, 99, 99, 101, 115, 115, 105, 111, 110).

Этот способ адресации назовем „естественно-языковая адресация”.

Учитывая, что в естественном языке слова имеют разную длину, а некоторые словосочетания являются понятиями, возникает требование возможности одновременной работы с взаимосвязанными



информационными пространствами различной размерности. Такую возможность предоставляет „Мультидоменная информационная модель” [Markov, 2004] и соответствующее программное обеспечение, названное - „Мультидоменный метод доступа” [Markov, 1984].

Если применить эту возможность к WordNet базе данных, получим результат более понятный для человека (Рис.8. б) и в тоже самое время – полностью понимаемый компьютером.

```
13082910 21 n 02 accession 0 addition 0 001 @ 13082742 n 0000 |
something added to what you already have; "the librarian
shelved the new accessions"; "he was a new addition to the
staff"
```

a)

```
accession 21 n 02 ; 0 addition 0 001 @ acquisition n 0000 |
something added to what you already have; "the librarian
shelved the new accessions"; "he was a new addition to the
staff"
```

б)

Рис. 8 Синсет слова „accession” из (а) настоящей и (б) NL-версии WordNet файла для существительных

## Заключение

В настоящей работе была представлена идея естественно-языковой адресации. Это дополнительная возможность для представления онтологической информации в интеллектуальных системах. Она имеет ряд преимуществ. На первом месте это линейная алгоритмическая сложность, которая зависит от максимальной длины слов ( $\max\_L$ ), а не от их количества, т.е.  $O(\max\_L)$ . Во-вторых, это уменьшение объема занимаемой памяти из-за полного отсутствия дополнительных индексов, абсолютных адресов и дополнительных файлов. Можно указать на уменьшение времени обработки вследствие полного отсутствия поиска – информация извлекается по прямому адресу. И не на последнем месте – универсальное представление информации одновременно доступной как для человека, так и для автоматизированных систем.

Этот способ организации информации используется в „Инструментальном комплексе онтологического назначения” (ИКОН) [Palagin et al, 2011], который разрабатывается в Институте кибернетики им. В.М.Глушкова НАНУ, Киев. Естественно-языковая адресация предусмотрена для хранения и использования информации в библиотеках онтологий, терминов и понятий, текстовых документов.

## Благодарности

Работа опубликована при финансовой поддержке проекта **ITHEA XXI** Института информационных теорий и приложений FOI ITHEA Болгария [www.ithea.org](http://www.ithea.org) и Ассоциации создателей и пользователей интеллектуальных систем ADUIS Украина [www.aduis.com.ua](http://www.aduis.com.ua).

## Библиография

[Markov, 1984] Kr.Markov. A Multi-domain Access Method. // Proceedings of the International Conference on Computer Based Scientific Research. Plovdiv, 1984. pp. 558-563.

[Markov, 2004] Markov, K. Multi-domain information model. Int. J. Information Theories and Applications, 11/4, 2004, pp.303-308.

[Palagin et al, 2011] А.В. Палагин, С.Л. Кривый, Н.Г. Петренко. Онтологические методы и средства обработки предметных знаний: монография/Луганск: изд-во ВЛУ им. В. Даля, 2011. – 323 с.

[Prutkov, 1863] Прутков Козьма Петрович: Сочинения. 1863г. [Электронный ресурс] // URL: [http://az.lib.ru/p/prutkow\\_k\\_p/](http://az.lib.ru/p/prutkow_k_p/) (дата обращения 04.08.2012).

[sw, 2012] Semantic Web: [Электронный ресурс] // URL: <http://www.w3.org/2001/sw/> (дата обращения 04.08.2012).

[WordNet, 2012] WordNet®. A lexical database for English. Princeton University. [Электронный ресурс] // URL: <http://wordnet.princeton.edu/> (дата обращения 04.08.2012).

---

### Информация об авторах

---



**Krassimira Ivanova** – *University of National and World Economy, Sofia, Bulgaria*

*e-mail: krazy78@mail.bg*

*Major Fields of Scientific Research: Data Mining*



**Vitalii Velychko** – *Institute of Cybernetics, NASU, Kiev, Ukraine*

*e-mail: velychko@aduis.com.ua*

*Major Fields of Scientific Research: Data Mining, Natural Language Processing*



**Krassimir Markov** – *Institute of Mathematics and Informatics at BAS, Sofia, Bulgaria;*

*e-mail: markov@foibg.com*

*Major Fields of Scientific Research: Multi-dimensional information systems, Data Mining*



## ITHEA® International Scientific Society

To extend the possibilities for international scientific collaboration in all directions of informatics by wide range of concrete activities, in 2002 year, the Institute for Information Theories and Applications FOI ITHEA® (IITA FOI ITHEA®) has been established as an international nongovernmental organization. IITA FOI ITHEA® is aimed to support international scientific research through international scientific projects, workshops, conferences, journals, book series, etc. The achieved results are remarkable. IITA FOI ITHEA® became world-wide known scientific organization. One of the main activities of the IITA FOI ITHEA® is building the ITHEA® International Scientific Society aimed to unite researches from all over the world who are working in the area of informatics.

Till the beginning of 2012 year, the **ITHEA® International Scientific Society** was joined by more than **2800** members from **48** countries all over the world: *Armenia, Azerbaijan, Belarus, Brazil, Belgium, Bulgaria, Canada, China, Czech Republic, Denmark, Egypt, Estonia, Finland, France, Germany, Greece, Hungary, India, Iran, Ireland, Israel, Italy, Japan, Jordan, Kyrgyz Republic, Latvia, Lithuania, Malaysia, Malta, Mexico, Moldova, Netherlands, Peru, Poland, Portugal, Romania, Russia, Scotland, Senegal, Serbia, Montenegro, Sweden, Spain, Sultanate of Oman, Turkey, UK, Ukraine, USA.*

ITHEA® Publishing House (**ITHEA® PH**) is the official publisher of the works of the ITHEA® ISS. The scope of the books of the ITHEA® ISS covers the area of Informatics and Computer Science. ITHEA® PH welcomes scientific papers and books connected with any information theory or its application.

ITHEA® ISS has two International Journals, established as independent scientific printed and electronic media, published by ITHEA® Publishing House:

- **International Journal “Information Theories and Applications” (IJ ITA)**, since 1993
- **International Journal “Information Technologies and Knowledge” (IJ ITK)** since 2007.

All journals and books of ITHEA® ISS are edited by the Institute of Information Theories and Applications FOI ITHEA in collaboration with the leading researchers from the: *Institute of Cybernetics “V.M.Glushkov”, NASU (Ukraine); Institute of Mathematics and Informatics, BAS (Bulgaria); University of Hasselt (Belgium); Natural Computing Group (NCG) of the Technical University of Madrid (Spain); Astrakhan State Technical University (Russia); Taras Shevchenko National University of Kiev (Ukraine); University of Calgary (Canada); VLSI Systems Centre, Ben-Gurion University (Israel).*

The great success of ITHEA® International Journals, Books and Conferences belongs to the whole of the ITHEA® International Scientific Society.

More information may be obtained from [www.ithea.org](http://www.ithea.org) .



Научное издание

Сборник статей

«ПРОБЛЕМЫ  
ИНТЕЛЛЕКТУАЛИЗАЦИИ  
КОМПЬЮТЕРА»

На русском и английском языках

Компьютерная верстка: В.Ю. Величко

Подп. в печать 03.09.2012. Формат 60x84/8. Бумага офс. Печать цифровая. Усл. печ. л. 43,71. Усл. краско-оттиск. 44,63. Уч.-изд. л. 47,0. Зак. 94. Тираж 60 экз.

Напечатано в авторской редакции с оригинал-макета заказчика на полиграфическом участке редакционно-издательского отдела Института кибернетики имени В.М. Глушкова НАН Украины 03680, ГСП, Киев-187, проспект Академика Глушкова,40



---

**ISBN: 978-966-02-6529-5 (printed)**  
**ISBN: 978-954-16-0061-0 (printed)**

---