

Хорошевский В.Ф.
(ВЦ РАН, Москва)



**ОЦЕНКА СИСТЕМ ИЗВЛЕЧЕНИЯ ИНФОРМАЦИИ
ИЗ ТЕКСТОВ НА ЕСТЕСТВЕННОМ ЯЗЫКЕ:
КТО ВИНОВАТ, ЧТО ДЕЛАТЬ**

План презентации

- Введение
- Цели и задачи работы
- Краткий обзор существующего положения
 - ◆ Проекты извлечения информации из текстов
 - ◆ Существующие метрики оценки систем типа IE
- Предлагаемые метрики оценки систем типа IE
 - ◆ Основные требования к системе метрик
 - ◆ Метрики оценки качества обработки объектов
 - ◆ Метрики оценки качества обработки отношений
 - ◆ Интегральные оценки качества систем типа IE
- Тестирование новой системы метрик
- Полученные результаты и дальнейшие исследования

Введение

- ❑ **Автоматическая обработка текстов на естественном языке (ЕЯ) - одно из приоритетных направлений в области новых информационных технологий.**
- ❑ **На современном этапе основное внимание исследователей и разработчиков практически значимых ЕЯ-систем сосредоточено на системах типа Multilingual Information Extraction (MIE), Summarization (Sum) и Question-Answering (QA).**

=====
=====

При этом одной из важнейших задач является оценка качества функционирования таких систем, без чего невозможны ни понимание достигнутых результатов, ни корректное сравнение разных систем.

Проект OntosMiner

реализуется
швейцарской фирмой Ontos AG и российской IT-фирмой АвиКомп

Мотивация разработки:

- ❑ Создание многоплатформенного семейства мультязыковых систем извлечения информации из ЕЯ-текстов.
- ❑ Представление результатов обработки в виде когнитивных карт.
- ❑ Использование результатов обработки в интеллектуальных аналитических системах.
- ❑ Использование результатов обработки в интеллектуальных системах навигации и поиска.

Проект OntosMiner

Текущее состояние проекта:

- ❑ Разработан инструментарий для создания систем типа IE на базе среды GATE.
 - ❑ Разработана среда LightOntos, обеспечивающая технологический цикл обработки текстов.
-
- ❑ Разработана система OntosMiner/English 3.0 для предметной области «Бизнес: Люди и Компании».
 - ❑ Разработана система OntosMiner/German для предметной области «Полицейские сводки».
 - ❑ Разработан прототип системы OntosMiner/French 1.0 для предметной области «Бизнес: Люди и Компании».
 - ❑ Разработано семейство систем OntosMiner/Russian для различных предметных областей.

Введение

Проект OntosMiner

The screenshot displays a desktop environment with two windows. On the left, a Notepad window titled 'Google.txt - Notepad' contains text about Google's management team and founders. On the right, a 'GoogleNew' browser window shows a complex network graph visualization. The graph consists of numerous nodes, many of which are highlighted in yellow, connected by a dense web of black lines. The nodes are arranged in a radial pattern, with a central hub and many smaller clusters branching out. The browser window's address bar and menu bar are visible at the top.

Google Inc. Management Team

Dr. Eric E. Schmidt
Chairman of the Board and Chief Executive Officer

Google founders Larry Page and Sergey Brin focused on strategic planning, management and technology, ensuring that quality remains high while focused on building the corporate infrastructure. Brin, Schmidt shares responsibility for Google's 20-year record of achievement as an Internet engine with a unique corporate culture.

Prior to his appointment at Novell, Schmidt led the Computer Science Lab at Xerox Palo Alto. Schmidt has a bachelor of science and Ph.D. in computer science from the University of Michigan.

Sergey Brin
Co-Founder & President, Technology

Sergey Brin, a native of Moscow, received his M.S. in computer science from the University of Maryland and his Ph.D. in computer science at Stanford University. Brin was a Science Foundation Graduate Fellow at Stanford where he met Larry Page and worked on the development of the PageRank algorithm. Brin and Brin continues to share responsibility for Google's success.

Brin's research interests include search engines, large text collections and scientific data analysis. His research includes Extracting Patterns and Relations from Large Text Collections, High Dimensionality, which he published in the Journal of Dynamic Itemset Counting and Implications, Association Rules to Correlations.

Brin has been a featured speaker at several international forums, including the World Economic Forum and the Technology Industry and the future of the Internet.

Введение

Проект OntosMiner

The screenshot displays the '6-01 [RussianIndexer] - LightOntos™ 2.0 for Workgroups' application. The interface includes a 'Management Console' on the left with a tree view of 'Knowledge Bases' (RussianIndexer) containing folders for 'Scheme', 'Facts', 'Queries', 'Fact Diagrams', 'Folders', and 'Search Queries'. The main workspace shows a knowledge graph with a central node 'Фрагмент_документа_282' connected to various other nodes via relationships like 'По документу_проход' and 'Дислокация'. The nodes include names and locations such as 'Сергеевом Посаде Московской обл.', 'ООО «Мастерок»', 'Ногинского филиала', 'Сухова-Кобылкин Максим Илларионович', 'Министерства здравоохранения РФ', 'станция технического обслуживания автомобилей «Volvo Lada»', 'Переделькина (Перестройкина) Анна Ивановна', 'с. Николаевка Сельскохозяйственного района Павлодарской области Казахской ССР', 'Кольянов Рашид Абдрахманович', 'с. Михайловское Псковской обл.', and 'ТОО «Стройсервис-Э.Ю.Я.»'. A text fragment at the bottom of the window contains the following text:

Участники Всероссийского конкурса «Чемелье руки», 6. Михайловское Псковской обл., 2006 г.:
- Кольянов Рашид Абдрахманович, 06.10.1974 г.р., уроженец с. Николаевка Сельскохозяйственного района Павлодарской области Казахской ССР. Работает старшим ТОО «Стройсервис-Э.Ю.Я.» Министерства здравоохранения РФ.
- Сухова-Кобылкин Максим Илларионович, 22.04.1970 г.р., каменщик Ногинского филиала (Московская обл., Ногинский район, г. Ногинск, ул. Строителей, 22) ООО «Мастерок», зарегистрированного в Сергеевом Посаде Московской обл. по адресу: ул. Подмосковная д.12;
- Переделькина (Перестройкина) Анна Ивановна, 15.02.1964 г.р., Работает мастером станции технического обслуживания автомобилей «Volvo & Lada».

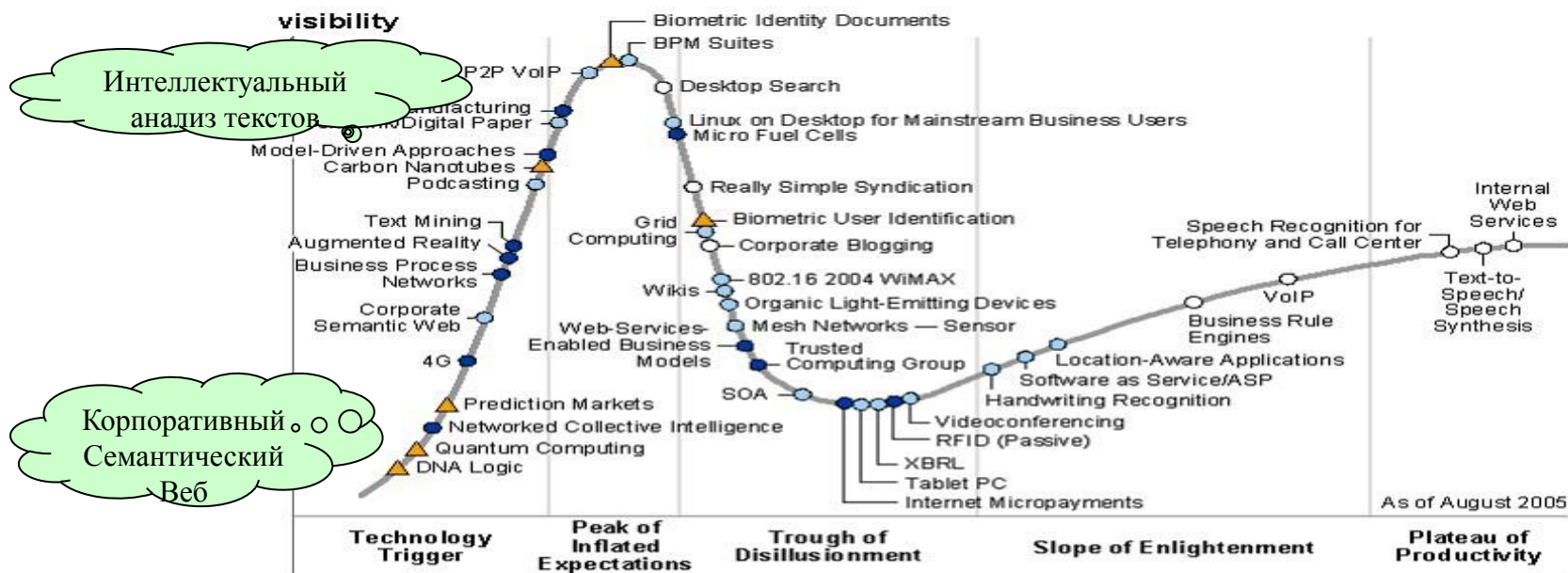
Diagram of instances: "6-01"

Цели и задачи настоящей работы

- ❑ **Обсуждение метрик для оценки качества функционирования систем извлечения информации из текстов на естественном языке**
- ❑ **Предложения по новой системе метрик.**
- ❑ **Анализ опыта использования новой системы метрик для оценки качества IE-систем, разрабатываемых в рамках проекта OntosMiner.**

Краткий обзор существующего положения

Новые технологии и продукты (Прогноз Gartner Group 2005)



Интеллектуальный анализ текстов

Корпоративный Семантический Веб

Plateau will be reached in:
 ○ less than 2 years ● 2 to 5 years ◆ 5 to 10 years ▲ more than 10 years ⊗ obsolete before plateau

Acronym Key

4G	fourth generation	SOA	service-oriented architecture
ASP	application service provider	VoIP	voice over Internet Protocol
BPM	business process management	WiMAX	Worldwide Interoperability for Microwave Access
P2P	peer to peer	XBRL	Extensible Business Reporting Language
RFID	radio frequency identification		

Проекты извлечения информации из текстов

□ США

- ◆ Проект TIPSTER (1991 – 1998)
 - Конференции TREC, MUC
- ◆ Программа TIDES (1999 – настоящее время)
 - Конференции MUC, DUC

□ Европа

- ◆ Форум CLEF
 - Конференции CLIF

□ Япония

- ◆ Форум NTCIR
 - Конференции NTCIR

□ Россия

- ◆ Инициатива РОМИП (2003 – настоящее время)
 - Семинары РОМИП

Проекты извлечения информации из текстов

Проект TIPSTER – новые технологии обработки текстов (1991 – 1998).

Управление проектом: DARPA, DoD, CIA в сотрудничестве с NIST.

Бюджет проекта: сотни млн. долл.

Основные направления:

- ❑ **Обнаружение специфической информации в тексте и извлечение ее из текста.**
- ❑ **Обнаружение документов, содержащих информацию, которая нужна пользователю (из потока текстов и\или хранилищ документов).**
- ❑ **Получение реферата документа или коллекции документов.**

Проекты извлечения информации из текстов

Программа TIDES (Translingual Information Detection, Extraction, and Summarization) – новая программа DARPA, специально ориентированная на получение новых результатов в технологиях реферирования документов и коллекций документов (1999 – настоящее время).

Основные направления:

- ❑ Извлечение информации из многоязыковых коллекций документов.
- ❑ Получение реферата/аннотации документа или коллекции документов.
- ❑ Разработка вопрос-ответных систем.

Проекты извлечения информации из текстов


Европа: Форум CLEF

Welcome to Cross Language Evaluation Forum - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://clef.isti.cnr.it/>

Cross Language Evaluation Forum



- Home
- Coordination
- CLEF 2006
- CLEF 2005
- CLEF 2004
- CLEF 2003
- CLEF 2002
- CLEF 2001
- CLEF 2000
- Publications
- Links
- Archives
- Contact
- Steering Committee
- Sponsors

The Cross-Language Evaluation Forum (CLEF) supports global digital library applications by (i) developing an infrastructure for the testing, tuning and evaluation of information retrieval systems operating on European languages in both monolingual and cross-language contexts, and (ii) creating test-suites of reusable data which can be employed by system developers for benchmarking purposes.


Through the organisation of system evaluation campaigns, the aim is to create a community of researchers and developers studying the same problems and to facilitate future collaborative initiatives between groups with similar interests. CLEF will also establish strong links, exchanging ideas and sharing results, with similar cross-language evaluation initiatives in the US and Asia, working on other sets of languages. The final goal is to assist and stimulate the development of European cross-language retrieval systems in order to guarantee their competitiveness on the global marketplace.

Registration for CLEF 2006 is now open.

The results of the CLEF2005 campaign were presented at the CLEF 2005 Workshop, 21 - 23 September, Vienna, Austria.

The CLEF 2005 Working Notes are accessible online.

CLEF is an activity of the [DELOS Network of Excellence for Digital Libraries](#) under the Sixth Framework Programme of the European Commission



Webdesign by Centromedia.com

Проекты извлечения информации из текстов

Япония: Форум NTCIR

The screenshot shows a Microsoft Internet Explorer browser window displaying the NTCIR Project Overview page. The browser's address bar shows the URL: <http://research.nii.ac.jp/ntcir/outline/prop-en.html>. The page title is "NTCIR Project Overview - Microsoft Internet Explorer". The browser's menu bar includes File, Edit, View, Favorites, Tools, and Help. The toolbar contains various navigation and utility icons. The page content features the NTCIR logo, a search box, and a navigation menu on the left. The main content area includes a title "NTCIR Project Overview" and a paragraph describing the NTCIR Workshop.

NTCIR (NII Test Collection for IR Systems) Project | [Related URL's](#) | [Contact Information](#) | [NII](#) |

NTCIR

Search

Search

[HOME](#)

[About NTCIR](#)

[WorkShop](#)

- [NTCIR-6](#)
- [NTCIR-5](#)
- [NTCIR-4](#)
- [NTCIR-3](#)
- [NTCIR-2](#)
- [NTCIR-1](#)

[Data](#)

[Overview](#) | [Original Proposal](#) |

NTCIR Project Overview

[[Japanese](#)]

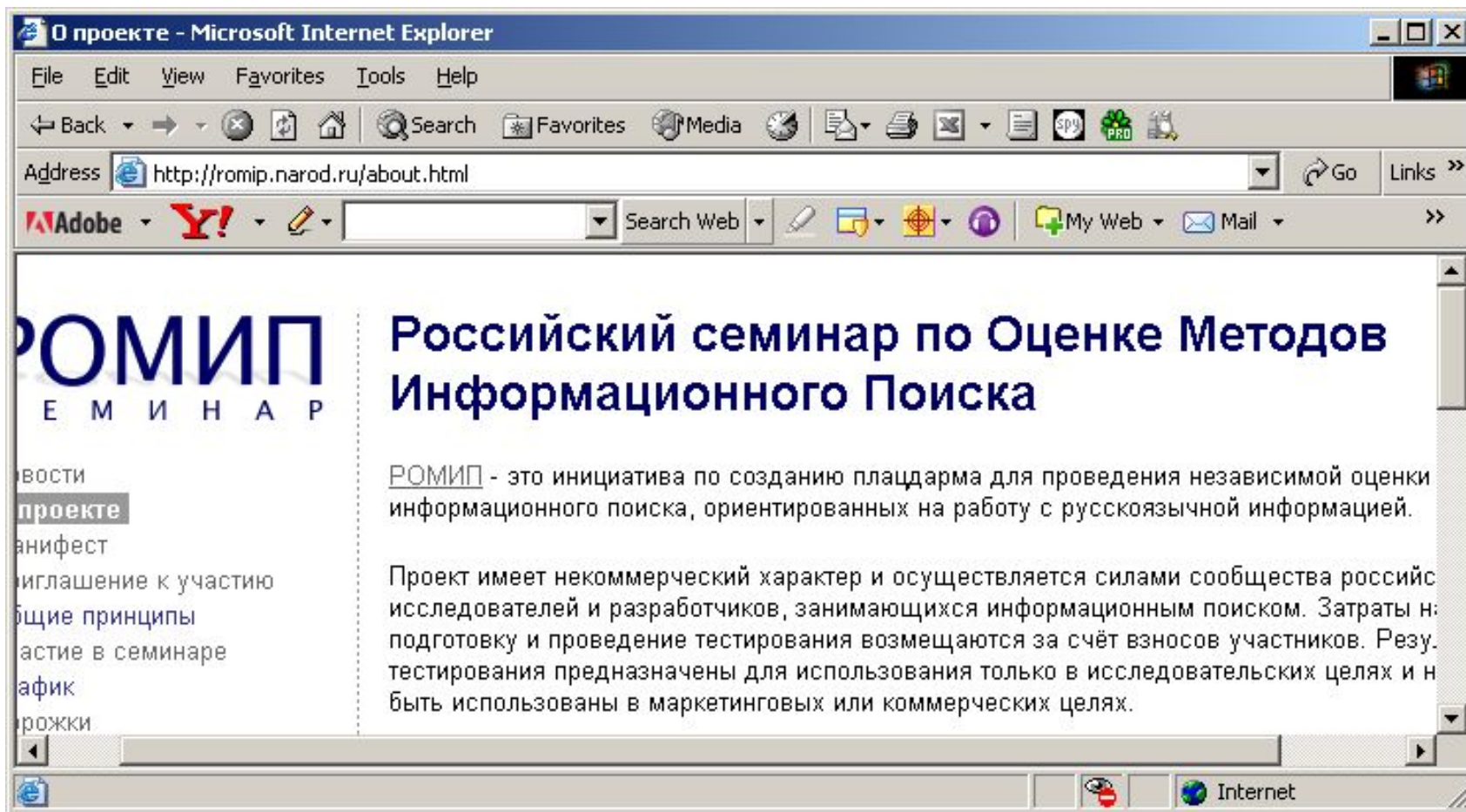
The NTCIR Workshop is a series of evaluation workshops designed to enhance research in Information Access (IA) technologies including information retrieval, question answering, text summarization, extraction, etc. It was co-sponsored by Japan Society for Promotion of Science (JSPS) as part of JSPS "Research for Future" Program" and National Center for Science Information Systems (NACSIS) since 1997, by JSPS and Research Center for Information Resources at National Institute of Informatics (RCIR/NII,) in FY 2000, and by MEXT Grant-in-Aid for Scientific Research on Priority Areas of "Informatics" (#13224087) and RCIR/NII in and after FY2001.

The aims are;

1. to encourage research in Information Access technologies by providing large-

Проекты извлечения информации из текстов

Россия: Инициатива РОМИП



The screenshot shows a Microsoft Internet Explorer browser window. The title bar reads "0 проекте - Microsoft Internet Explorer". The address bar contains "http://romip.narod.ru/about.html". The page content includes a logo for "РОМИП" (ROMIP) with the letters "Е М И Н А Р" below it. The main heading is "Российский семинар по Оценке Методов Информационного Поиска" (Russian Seminar on Evaluation of Information Search Methods). The text describes the initiative as a platform for independent evaluation of information search methods, aimed at working with Russian-language information. It mentions that the project is non-commercial and supported by a community of researchers and developers. The seminar is intended for research purposes only and will not be used for marketing or commercial goals.

РОМИП
Е М И Н А Р

Российский семинар по Оценке Методов Информационного Поиска

РОМИП - это инициатива по созданию плацдарма для проведения независимой оценки информационного поиска, ориентированных на работу с русскоязычной информацией.

Проект имеет некоммерческий характер и осуществляется силами сообщества российских исследователей и разработчиков, занимающихся информационным поиском. Затраты на подготовку и проведение тестирования возмещаются за счёт взносов участников. Результаты тестирования предназначены для использования только в исследовательских целях и не будут использованы в маркетинговых или коммерческих целях.

Существующие метрики оценки систем типа IE

- ❑ Начало работам по метрикам оценки систем извлечения информации из текстов было положено исследованиями, которые проводились в 1991-1998 г.г. в рамках проекта TIPSTER.
- ❑ Хотя первые результаты в этом направлении были получены еще в классических работах Солтона по оценке систем информационного поиска.
- ❑ И, тем не менее, применительно к задаче оценки систем типа IE, одним из важных результатов проекта TIPSTER было создание соответствующего стандарта de Facto.
- ❑ Развитие работ по данному направлению в США происходило и происходит в рамках программы TIDES.
- ❑ В Великобритании акцент аналогичных работ на метриках, учитывающих использование онтологий.

В настоящее время исследования и разработки по метрикам для оценки систем типа IE ведутся не только в США, Великобритании, но и в континентальной Европе, Японии и России, но сосредоточены эти работы, в основном, на задачах оценки систем информационного поиска, кластеризации и вопрос-ответных систем.

Существующие метрики оценки систем типа IE

Основные требования:

- ❑ Значения метрик должны быть максимальными для «хороших» систем и минимальными для «плохих», а их изменение должно быть монотонно.
- ❑ Метрики должны быть понятными и интуитивно прозрачными, эффективно вычисляться и коррелировать с оценками эксперта-человека.
- ❑ Метрики не должны допускать разные варианты интерпретации результатов.

Метрики проекта TIPSTER и программы TIDES:

$$P = \frac{Correct + 1/2Partial}{Correct + Spurious + Partial}$$

$$R = \frac{Correct + 1/2Partial}{Correct + Miss + Partial}$$

$$F - measure = \frac{(\beta^2 + 1)P * R}{\beta^2 R + P}$$

- **точность** (P) - к-во правильно идентифицированных объектов как процент от общего количества идентифицированных объектов,
- **полнота** (R) - к-во правильно идентифицированных объектов как процент от общего количества правильных объектов,
- **качество** ($F-measure$) - интегральная метрика, которая является взвешенной оценкой параметров точности и полноты.

Существующие метрики оценки систем типа IE

Достоинства:

- Отвечают основному критерию, указанному выше и эффективно вычислимы, а также понятны эксперту.

Недостатки:

- Не всегда коррелируют с мнением экспертов и допускают различные интерпретации результатов.
- Ограниченность и неполнота. Оценивается качество выделения из текстов NE (например, объектов типа Person, Organization, Location и др.), но при этом не учитывают точность и полноту выделения артефактов, относящихся к этим NE (например, атрибутов типа JobTitle, Time и др.).
- Нет хороших метрик для оценки качества выделения отношений между выделенными объектами, а если такие метрики используются, то они, по существу, «штрафуют» систему несколько раз за одну и ту же ошибку.
- Не учитывается значимость компонент выделенных объектов и отношений в тех случаях, когда те имеют внутреннюю структуру.

Нужна новая система метрик!!!

Предлагаемые метрики оценки систем типа IE

Основные требования к системе метрик:

- ❑ **Монотонность всех метрик и системы метрик в целом.**
- ❑ **Сбалансированность всех метрик системы.**
- ❑ **Понятность и интуитивная прозрачность отдельных метрик и системы метрик в целом для эксперта-человека.**
- ❑ **Однозначность интерпретации результатов оценки.**
- ❑ **Возможность интегральной оценки качества.**
- ❑ **Эффективная вычислимость всех метрик и системы метрик в целом.**
- ❑ **Адекватность текущему уровню теории и практики создания IE-систем и возможность обобщения в будущем.**

Предлагаемые метрики оценки систем типа IE

Параметры новой системы метрик

Для существующих метрик оценки качества систем типа IE параметры их опираются на аннотации объектов (NEs). В качестве модели аннотаций в настоящее время, как правило, используется подход, принятый в проекте TIPSTER.

=====

Представляется, что такой подход может с успехом использоваться и в новой системе метрик.

=====

Будем полагать, что аннотация представляется в формате, где явно специфицированы тип выделенного объекта (отношения) и его атрибуты, а также расположение аннотации в тексте относительно его начала (Offsets).

Тогда общая структура аннотации – следующая:

```
<AnnName is_a AnnType; StartOffset = Number; EndOffset = Number;  
Attr1 = Value1; ... Attrn = Valuen >
```

Для простоты будем считать, что значениями атрибутов могут быть элементарные типы данных (например, string, integer и т.п.) или их одномерные массивы.

Метрики оценки качества обработки объектов

Оценка точности выделения объектов

Правильно идентифицированным будем называть такой объект, который, по мнению эксперта, зафиксирован в тексте правильно (правильны значения типа объекта и его OffSets) и все существенные атрибуты объекта тоже правильны (значения OffSets правильны, а имена атрибутов заполнены правильными значениями).

Таким образом,

$$Correct(X) = \begin{cases} 1, & \text{iff } Type(X) = true \ \& \ StartOffset(X) = true \ \& \ EndOffset(X) = true \ \& \\ & (\forall i, i = 1 \div k, Correct(Attr_i(X)) = 1) \\ 0, & \text{otherwise} \end{cases} .$$

Метрики оценки качества обработки объектов

Оценка точности выделения объектов

Тогда, по аналогии, полностью неправильно идентифицированный объект фиксируется следующей формулой:

$$Spurious(X) = \begin{cases} 1, \text{ iff } Type(X) = false \ \& \ StartOffset(X) = false \ \& \ EndOffset(X) = false \ \& \\ (\forall i, i = 1 \div k, Spurious(Attr_i(X)) = 1) \\ 0, \text{ otherwise} \end{cases} .$$

Метрики оценки качества обработки объектов

Оценка точности выделения объектов

Иначе обстоит дело с частично правильно идентифицированными объектами, так как в этом случае необходимо не только фиксировать данный феномен, но и правильно оценить «тяжесть» допущенных ошибок.

Понятно, что объект X можно рассматривать как *Partial* лишь в том случае, если $Type(X) = true$. Тогда для фиксации феномена целесообразно использовать формулу вида:

$$Partial(X) = \begin{cases} 1 \text{ with weight } \alpha(X), \text{ iff } Type(X) = true \ \& \\ \quad (StartOffset(X) = false \vee EndOffset(X) = false) \vee \\ \quad (\exists i, i = 1 \div k, Spurious(Attr_i(X)) = 1) \\ 0, \text{ otherwise} \end{cases}$$

где

$$\alpha(X) = \frac{\varepsilon(X) + \delta(X)}{2}$$

Метрики оценки качества обработки объектов

Оценка точности выделения объектов

$$\varepsilon(X) = \frac{1 - \sum_{i=1}^k \gamma_i * Spurious(Attr_i(X)) + \sum_{i=1}^k \gamma_i * Correct(Attr_i(X))}{2};$$

$$\delta(X) = \frac{1 - \delta_{Spurious}(X) + \delta_{Correct}(X)}{2};$$

$$\delta_{Spurious}(X) = \omega_{StartOffset} * Spurious(StartOffset(X)) + \omega_{EndOffset} * Spurious(EndOffset(X));$$

$$\delta_{Correct}(X) = \omega_{StartOffset} * Correct(StartOffset(X)) + \omega_{EndOffset} * Correct(EndOffset(X)).$$

где

- $\alpha(X)$, $\varepsilon(X)$ и $\delta(X)$ – коэффициенты качества обработки всего объекта, атрибутов объекта и качества фиксации позиций объекта в тексте соответственно, а
- γ и ω – веса атрибутов и Offsets, причем $0 \leq \gamma_i \leq 1$ и
- $\sum \gamma_i = 1$; $0 \leq \omega_{StartOffset} \leq 1$; $0 \leq \omega_{EndOffset} \leq 1$ и
- $(\omega_{StartOffset} + \omega_{EndOffset}) = 1$.

Метрики оценки качества обработки объектов

Оценка точности выделения объектов

С учетом введенных выше понятий

$$P_{NE} = \frac{\textit{Correct} + \sum_{X \in NE} \alpha(X) * \textit{Partial}(X)}{\textit{Correct} + \textit{Spurious} + \textit{Partial}}.$$

Метрики оценки качества обработки отношений

Общие замечания

- Оценки точности выделения отношений, в отличие от объектов, в настоящее время практически не имеют устойчивых метрик.
- На наш взгляд такая ситуация связана с тем, что
 - ◆ сами отношения лишь недавно стали обрабатываться в системах типа IE,
 - ◆ сложность их оценки значительно выше, чем сложность оценки объектов.

Ограничения на метрики для оценки качества выделения отношений:

- В силу того, что эксперты устойчиво фиксируют наличие отношений, но расходятся во мнении об экспликации их позиций в тексте, исключим из соответствующих метрик параметры OffSets (заметим, что часто отношение «размыто» в рамках предложения и даже всего текста уже по самой природе ЕЯ).
- При оценке качества выделения отношений необходимо учесть эффекты наведенных ошибок, связанных с неверной обработкой объектов и\или атрибутов, которые специфицируют его актаны (нельзя «наказывать» за одну и ту же ошибку несколько раз).
- Так как качество выделения отношения зависит не только от качества выделения актантов, но и от качества его собственной обработки, необходимо в соответствующих метриках явно оценивать и ту, и другую составляющие.

Метрики оценки качества обработки отношений

Оценка точности выделения отношений

Правильно идентифицированным будем называть такое **отношение**, которое, по мнению эксперта, присутствует в тексте и выделено системой, причем тип отношения между правильно выделенными обязательными актантами определен правильно и все существенные атрибуты отношения тоже правильны.

Т.е.

$$Correct(R) = \begin{cases} 1, \text{ iff } Type(R) = true \ \& \ (\forall j, j = 1 \div n, Correct(Act_j(R)) = 1) \ \& \\ \quad (\forall i, i = 1 \div k, Correct(Attr_i(R)) = 1) \\ 0, \text{ otherwise} \end{cases} .$$

Метрики оценки качества обработки отношений

Оценка точности выделения отношений

Тогда полностью неправильно идентифицированное отношение можно фиксировать следующей формулой:

$$Spurious(R) = \begin{cases} 1, & \text{iff } Type(R) = false \\ 0, & \text{otherwise} \end{cases} .$$

Метрики оценки качества обработки отношений

Оценка точности выделения отношений

Для частично правильно идентифицированных отношений, как и в случае объектов, необходимо правильно оценить «тяжесть» допущенных ошибок. Отношение R естественно оценивать как *Partial*, только если $Type(R) = true$. Тогда для фиксации феномена на верхнем уровне можно использовать формулу вида:

$$Partial(R) = \begin{cases} 1 \text{ with weight } \tilde{\alpha}(R), \text{ iff } Type(R) = true \ \& \\ \quad ((\exists j, j = 1 \div n, Spurious(Act_j(R)) = 1 \vee Partial(Act_j(R)) = 1) \vee \\ \quad (\exists i, i = 1 \div k, Spurious(Attr_i(R)) = 1 \vee Partial(Attr_i(R)) = 1)) \\ 0, \text{ otherwise} \end{cases}$$
$$\tilde{\alpha}(R) = \frac{\tilde{\varepsilon}(R) + \tilde{\delta}(R)}{2} \quad \tilde{\varepsilon}(R) = \frac{1 - \tilde{\varepsilon}_{Spurious}(R) + \tilde{\varepsilon}_{Correct}(R) + \tilde{\varepsilon}_{Dummy}(R)}{2}$$
$$\tilde{\delta}(R) = \frac{1 - \sum_{i=1}^k \gamma_i * Spurious(Attr_i(R)) + \sum_{i=1}^k \gamma_i * Correct(Attr_i(R))}{2}.$$

Метрики оценки качества обработки отношений

Оценка точности выделения отношений

$$\tilde{\varepsilon}(R) = \frac{1 - \tilde{\varepsilon}_{Spurious}(R) + \tilde{\varepsilon}_{Correct}(R) + \tilde{\varepsilon}_{Dummy}(R)}{2}$$

$$\tilde{\varepsilon}_{Spurious}(R) = \sum_{j=1}^n \gamma_j * Spurious(Act_j(R)) \quad \left| \begin{array}{l} \\ \exists Correct(Act_j(R)) \end{array} \right. ;$$

$$\tilde{\varepsilon}_{Correct}(R) = \sum_{j=1}^n \gamma_j * Correct(Act_j(R));$$

$$\tilde{\varepsilon}_{Dummy}(R) = \sum_{j=1}^n \gamma_j * Dummy(Act_j(R)) \quad \left| \begin{array}{l} \\ \exists Correct(Act_j(R)) \end{array} \right. ;$$

$\tilde{\alpha}(R)$, $\tilde{\delta}(R)$, $\tilde{\varepsilon}(R)$ – коэффициенты качества обработки всего отношения, его атрибутов и качества выделения актантов отношения соответственно,
 γ – веса атрибутов, причем $0 \leq \gamma_i \leq 1$ и $\sum \gamma_i = 1$.

Метрики оценки качества обработки отношений

Оценка точности выделения отношений

С учетом введенных выше понятий

$$P_R = \frac{\textit{Correct} + \sum_{X \in R} \tilde{\alpha}(R) * \textit{Partial}(R)}{\textit{Correct} + \textit{Spurious} + \textit{Partial}}.$$

Метрики оценки качества обработки объектов/отношений

Оценка полноты выделения объектов/отношений

В оценках полноты в классическом варианте участвуют те же параметры, что и в оценках точности, за исключением параметра *Miss*, характеризующего количество пропущенных системой объектов и/или отношений. Поэтому, определив этот параметр, как

$$Miss(X) = \begin{cases} 1, \text{ iff } Type(X) = undef \\ 0, \text{ otherwise} \end{cases} \quad \text{и} \quad Miss(R) = \begin{cases} 1, \text{ iff } Type(R) = undef \\ 0, \text{ otherwise} \end{cases} ,$$

можно, по аналогии с оценками точности, предложить следующие оценки для полноты:

$$R_{NE} = \frac{Correct + \sum_{X \in NE} \alpha(X) * Partial(X)}{Correct + Miss + Partial} ,$$

$$R_R = \frac{Correct + \sum_{X \in R} \tilde{\alpha}(R) * Partial(R)}{Correct + Miss + Partial} .$$

Интегральные оценки качества систем типа IE

В предлагаемой системе метрик для точности и полноты введены по две оценки – для объектов и отношений.

Для получения интегральной оценки точности целесообразно взвесить P_{NE} и P_R коэффициентом их относительной важности β_p по аналогии с F-мерой:

$$P = \frac{(\beta_p^2 + 1)P_{NE} * P_R}{\beta_p^2 P_R + P_{NE}}.$$

Аналогично:

$$R = \frac{(\beta_r^2 + 1)R_{NE} * R_R}{\beta_r^2 R_R + R_{NE}}.$$

Тогда:

$$F - measure = \frac{(\beta^2 + 1)P * R}{\beta^2 R + P}.$$

Тестирование новой системы метрик

Для тестирования новой системы метрик была проведена оценка качества процессора OntosMiner/Russian. Для этого был сформирован корпус документов, которые могут быть отнесены к предметной области, поддерживаемой этим процессором. Документы брались с русскоязычных новостных сайтов Интернет. Общий объем корпуса – 1882 док.

Типичный пример фрагмента текста из данного корпуса:

REGNUM, 09.02.2006 09:54

Комитет по экологии Госдумы будет настаивать на приостановке деятельности ПО "Маяк" 8 февраля в Государственной Думе состоялось заседание комитета по экологии. Основной темой повестки дня стало обсуждение экологических проблем, связанных с деятельностью ПО "Маяк", по результатам поездки в Озёрск, сообщили корреспонденту ИА REGNUM в пресс-центре Уральского Межрегионального координационного совета партии "Единая Россия" ...

Справка: Федеральное государственное унитарное предприятие (ФПГУ) "Производственное объединение "Маяк" - предприятие ядерного оружейного комплекса России. Оно входит в состав Федерального агентства по атомной энергии Российской Федерации.

ПО "Маяк" - единственное в России предприятия по переработке отработанного ядерного топлива. На предприятии перерабатывают ядерные отходы с 1977 года. Комбинат обслуживает Кольскую, Нововоронежскую и Белоярскую атомные станции, а также перерабатывает ядерное топливо с атомных подводных лодок и из-за рубежа.

Тестирование новой системы метрик

Для оценки результатов было решено использовать объекты типа Person, JobTitle/Title, Organization и Location, а также отношения типа BeEmployeeOf и ConnectedWith.

Такой набор является практически стандартным для оценок на конференциях TREC/MUC/DUC и дает интересную информацию по сравнению классической и предложенной системы метрик.

Процедура оценки была организована следующим образом:

- 6 текстов, случайно выбранных из контрольного корпуса, тэгировали эксперты-лингвисты.
- Результаты ручного тегирования сравнивались с результатами обработки этих же текстов системой OntosMiner/Russian.
- Вычисления оценок производились по классическим формулам и формулам, предложенным в данной работе.
- При этом, для простоты, предполагалось, что точность и полнота имеют одинаковые веса ($\beta=1$), а веса атрибутов и Offsets, если они используются при вычислениях, тоже одинаковы ($\gamma=1/k$, где k – кол-во атрибутов, и $\omega_{StartOffset} = \omega_{EndOffset} = 1/2$).

Тестирование новой системы метрик

Классические оценки

Объект/ Параметр	Док	Именованные сущности			
		Person	JobTitle	Organization	Location
<i>Correct / Partial / Spurious / Miss</i>	1.	4 / 1 / 0 / 0	10 / 0 / 0 / 0	15 / 3 / 1 / 2	9 / 0 / 0 / 0
	2.	13 / 0 / 0 / 1	9 / 0 / 1 / 5	12 / 1 / 0 / 1	30 / 0 / 0 / 0
	3.	21 / 0 / 0 / 0	11 / 1 / 0 / 4	16 / 1 / 0 / 6	15 / 0 / 0 / 0
	4.	56 / 0 / 2 / 0	34 / 0 / 1 / 3	103 / 4 / 1 / 23	35 / 0 / 2 / 5
	5.	12 / 0 / 0 / 4	10 / 1 / 1 / 1	11 / 2 / 2 / 5	22 / 0 / 1 / 1
	6.	25 / 2 / 1 / 1	26 / 0 / 0 / 0	22 / 2 / 1 / 7	13 / 0 / 2 / 0
<i>Точность / Полнота / F-мера</i>	1.	0,90 / 0,90 / 0,90	1,00 / 1,00 / 1,00	0,89 / 0,82 / 0,85	1,00 / 1,00 / 1,00
	2.	1,00 / 0,93 / 0,96	0,90 / 0,64 / 0,75	0,96 / 0,89 / 0,92	1,00 / 1,00 / 1,00
	3.	1,00 / 1,00 / 1,00	0,94 / 0,72 / 0,82	0,97 / 0,76 / 0,85	1,00 / 1,00 / 1,00
	4.	0,97 / 1,00 / 0,98	0,97 / 0,92 / 0,94	0,97 / 0,81 / 0,88	0,95 / 0,88 / 0,91
	5.	1,00 / 0,75 / 0,86	0,87 / 0,87 / 0,87	0,80 / 0,67 / 0,73	0,96 / 0,96 / 0,96
	6.	0,93 / 0,93 / 0,93	1,00 / 1,00 / 1,00	0,92 / 0,74 / 0,82	0,87 / 1,00 / 0,93

Тестирование новой системы метрик

Предлагаемые оценки (объекты)

Объект/ Параметр	Док	Именованные сущности			
		Person	JobTitle	Organization	Location
<i>Correct / Partial / Spurious / Miss</i>	1.	4 / 1 / 0 / 0	10 / 0 / 0 / 0	15 / 3 / 1 / 2	9 / 0 / 0 / 0
	2.	13 / 0 / 0 / 1	9 / 0 / 1 / 5	11 / 1 / 1 / 1	30 / 0 / 0 / 0
	3.	21 / 0 / 0 / 0	11 / 1 / 0 / 4	16 / 1 / 0 / 6	15 / 0 / 0 / 0
<i>Точность / Полнота / F-мера</i>	1.	0,90 / 0,90 / 0,90	1,00 / 1,00 / 1,00	0,83 / 0,79 / 0,81	1,00 / 1,00 / 1,00
	2.	1,00 / 0,93 / 0,96	0,90 / 0,64 / 0,75	0,94 / 0,88 / 0,91	1,00 / 1,00 / 1,00
	3.	0,98 / 0,99 / 0,99	0,99 / 0,72 / 0,82	0,96 / 0,71 / 0,82	1,00 / 1,00 / 1,00

Тестирование новой системы метрик

Предлагаемые оценки (отношения)

Объект/ Параметр	Док	Отношения	
		<i>BeEmployeeOf</i>	<i>ConnectedWith</i>
<i>Correct / Partial / Spurious / Miss</i>	1.	2 / 0 / 0 / 2	24 / 0 / 0 / 0
	2.	4 / 0 / 0 / 1	21 / 0 / 0 / 0
	3.	2 / 0 / 0 / 3	35 / 0 / 0 / 0
<i>Точность / Полнота / F-мера</i>	1.	1,00 / 0,50 / 0,67	1,00 / 1,00 / 1,00
	2.	1,00 / 0,80 / 0,89	1,00 / 1,00 / 1,00
	3.	1,00 / 0,40 / 0,57	1,00 / 1,00 / 1,00

Полученные результаты и дальнейшие исследования

Анализ полученных результатов:

- ❑ Новые метрики более «чувствительны» к ошибкам в определении Offsets,
- ❑ Новые метрики учитывают важность атрибутов и правильность их выделения.
- ❑ Аналогичная ситуация имеет место и для оценок отношений.

Направления дальнейших исследований:

1. Необходим учет правильной и\или неправильной обработки орфосинонимических объектов.
2. Необходима разработка автоматической системы тестирования с удобными средствами настройки расчетных формул, развитыми средствами сбора статистики и визуализации полученных результатов.
3. Необходимо исследование поведения системы метрик в различных (в том числе пограничных) случаях.

Хорошевский В.Ф.



**БЛАГОДАРЮ
ЗА ВНИМАНИЕ**