

бизнес-сценарий AviComp Services

## **Семантические метаданные для веб-порталов**

оглавление

что такое семантические метаданные 3

как их создавать и использовать 9

AviCon

## что такое семантические метаданные

Семантические метаданные (англ. semantic metadata) — дополнительный слой информации, позволяющий автоматизировать установление связей между различными частями контента.

### — Во-первых, новые связи между частями контента

Самый простой пример — привязка дополнительных ссылок к новостному тексту на информационном портале. Обычно для этого используется полуавтоматическая кластеризация текстов (распределение текстов по различным темам). Недостатки этого подхода хорошо известны:

- нагрузка на персонал
- влияние человеческого фактора
- низкие объемы обрабатываемой информации и как следствие — временами низкая релевантность связанных ссылок и/или недостаточное их количество

Семантические метаданные позволяют обогатить имеющийся контент принципиально новыми связями, которые невозможно установить при стандартной кластеризации.

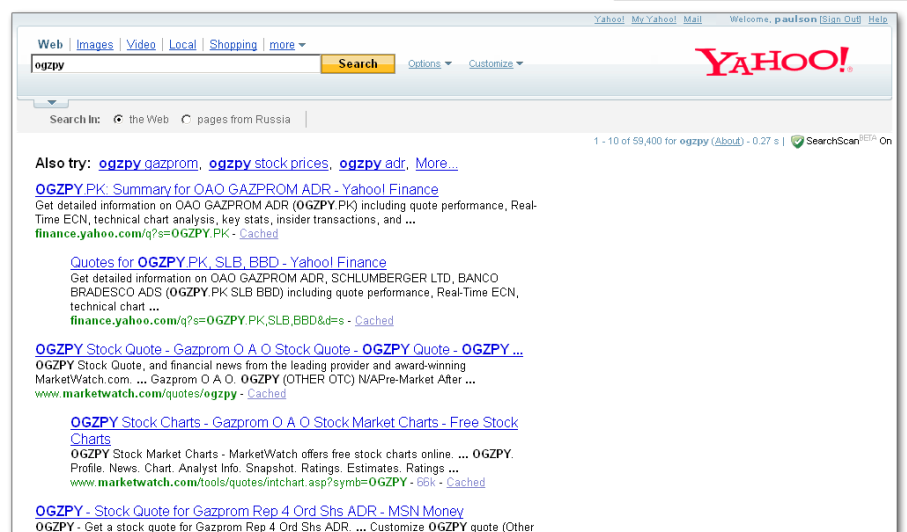
### — Во-вторых, принципиально новый подход к поиску

Сегодня мы настолько привыкли к поиску по ключевым словам, что уже забыли, насколько это неудобно. Пытаясь найти нужную информацию, пользователь изучает несколько десятков ссылок, не зная заранее,

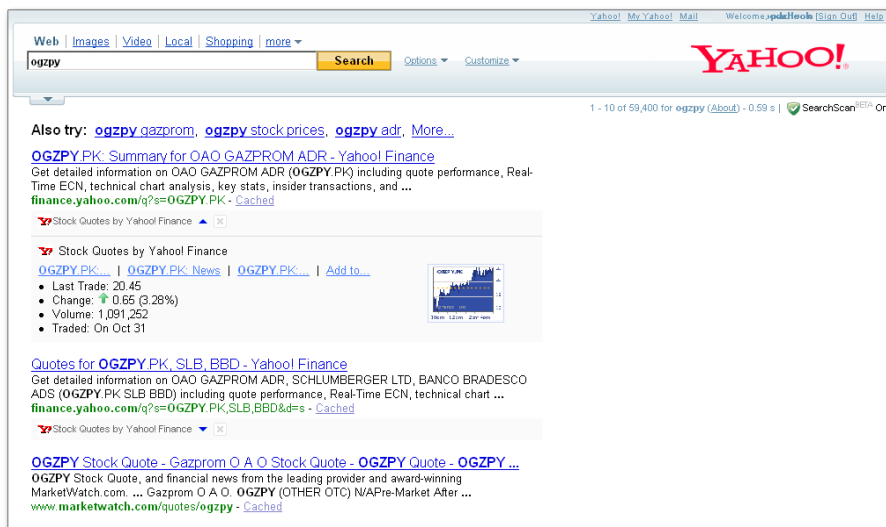
найдет ли он то, что ему нужно, или нет. Гораздо удобнее иметь поиск, выдающий однозначный результат, который содержит в структурированном виде всю информацию об искомом объекте.

Пример — поиск информации о ходе торгов бумагами Газпрома на бирже (тикер OGZPY).

До подключения семантических метаданных:



После:



В отличие от поисковых «шорткатов» (например, выдача данных о погоде в ответ на запрос «погода в Москве»), выдача семантического контента гораздо менее не ограничена тематикой и структурой запроса.

### — В-третьих, дополнительные возможности навигации

Семантические метаданные можно представить в виде метаразметки на страницах веб-ресурса. Почему бы эту метаразметку не сделать видимой пользователю?

Сегодня на информационных и развлекательных порталах практикуется создание в тексте гиперссылок на вхождения имен, компаний, географических названий и т.д. (редакторская разметка). Эти ссылки связывают объекты с отдельными страницами, где вручную собрана основная интересующая пользователя информация.

Семантические метаданные позволяют автоматически распознавать вхождения этих объектов и генерировать структурированный контент о данном объекте на основе индексируемых ресурсов. При помощи проиндексированных семантических связей между объектами пользователь получает возможность навигироваться по контенту вне зависимости от структуры веб-портала.

На следующих скриншотах семантическая навигация реализована в виде всплывающего мини-браузера, содержащего всю информацию о выбранном объекте, добытую на различных ресурсах.

AviCo

До подключения семантической навигации:

The screenshot shows the inoSMI.Ru website interface. At the top, there are navigation links for 'текстовая версия', 'Подписка', 'Экспорт', 'Форум', and 'Все за сегодня'. Below the search bar, there are several news teasers, including one from 'guardian.co.uk' titled 'Когда Михаил подался в религию ("The Guardian", Великобритания)'. The main article text discusses Gorbachev's religious beliefs and his visit to the tomb of St. Francis of Assisi. A yellow arrow points from a box labeled 'Редакторская разметка' to the text 'Михаил Горбачев - человек, при котором рухнул Советский Союз - дал понять, что он верующий.' The right sidebar contains a 'Новости Партии' section with various news items.

После:

Мини-браузер

Связи объекта

Персона

Семантическая разметка

Ссылки на релевантные статьи

inoСМИ.Ru

Последние переводы

В России  
В Мире  
Панорама  
Радио  
ИноВидео  
Аналитика  
Пресс-клуб  
ЭКСПО-2005

Актуальные сюжеты  
Заголовки дня  
О проекте

guardian.co.uk

Когда Михаил подался в религию ("...")  
В свои 77 лет Горбачев - самый пожилой из бывших лидеров СССР, проявивших интерес к Богу  
Люк Хардинг (Luke Harding), 20 марта 2008

Он был лидером последней в мире агентств...  
он был генеральным секретарем коммунистической партии. Но на этой неделе Михаил Горбачев - лидер СССР - дал понять, что он не верит в Бога.

Во время визита в Италию...  
Ассизского, где вместе со своей дочерью...  
размышлении. После этого он объявил, что...  
'фундаментальную' роль в его жизни...  
Франциска, поэтому мне было важно...  
после...

Высказывания Горбачева...  
частично и публичными...  
в жизни...  
в жизни...

Персона  
[Navigate]

Object relations:

упоминается | встречи\_и\_переговоры | связи\_с\_организациями | связи\_с\_другими\_персонами | работает | высказывания\_заявления | цитаты | поддерживает | отношения\_с\_полит\_кланнами | упоминает | упоминается\_с |

Mentioned in documents:

#	Title	Rank	Threshold
1.	ИноСМИ.Ru   Горбачев приветствует достижения Путина	100	100
2.	ИноСМИ.Ru   Победил рынок над демократией	100	100
3.	ИноСМИ.Ru   Горбачев поддерживает Путина на выборах и говорит, что Россия - серьезный партнер Запада	100	100
4.	Би-би-си   Россия   Горбачев поддержал Путина	100	100
5.	NR2.Ru: Горбачев поддерживает «Единую Россию» и Путина, но сочувствует Каспарову / 29.11.07 / Центральный федеральный округ	100	100
6.	NR2.Ru: Горбачев поддерживает «Единую Россию» и Путина, но сочувствует Каспарову / 29.11.07 / Центральный федеральный округ	100	100
7.	NR2.Ru: Горбачев поддерживает «Единую Россию» и Путина, но сочувствует Каспарову / 29.11.07 / Новый Регион - Политика	100	100
8.	NR2.Ru: Горбачев поддерживает «Единую Россию» и Путина, но сочувствует Каспарову / 29.11.07 / Новый Регион - Политика	100	100
9.	ИноСМИ.Ru   Когда Михаил подался в религию	100	100
10.	М. Горбачев: Избирательная система РФ нуждается в корректировке :: Политика :: Top-Info.ru	100	100

Дизайн навигации легко настраивается на общий вид портала (минибраузер интерпретирует имеющиеся CSS, также возможна ручная настройка). Для использования навигации необходимо лишь вставить небольшой скрипт в исходный код страницы.



Сервис семантической навигации может эксплуатировать минибраузер или виджеты, встроенные в макет страницы:

The screenshot displays a web interface for semantic navigation. On the left, a sidebar titled 'Объекты' (Objects) contains a search bar and a list of geographical locations with horizontal bar charts indicating their relevance. The main content area, titled 'Дмитрий Медведев' в деталях [Персона] (Dmitry Medvedev in details [Person]), shows a list of related items such as 'Назначения Увольнения (1 Item)', 'встречи и переговоры (14 Items)', and 'деятельность в контексте выборов (5 Items)'. A search results window at the bottom right shows a result for 'Дмитрий Медведев / работает / комитет по международным отношениям городской администрации' in the media, sorted by date, with a snippet from inosmi.ru dated July 7.

## как создавать и использовать семантические метаданные

Совместно со швейцарской компанией Ontos International мы разработали и запатентовали собственную технологию обработки неструктурированных данных и автоматического создания семантического контента. В основе нашей технологии — лингвистический процессор **OntosMiner™**, распознающий



объекты различных типов (имена людей, названия компаний, продукты, местоположения и т.д.) и связи между ними (положительные/отрицательные мнения об объекте, высказывания объекта и об объекте, «работать в», «располагается в», «лоббировать», «быть конкурентом» и т.д.).

Результаты обработки информационных ресурсов агрегируются в специализированном хранилище, причем по мере поступления новой информации дублирующиеся объекты идентифицируются и объединяются. Хранилище поддерживает стандарты, рекомендованные W3C (RDF, OWL, N3, Turtle и др.), что позволяет легко интегрировать наши решения с существующими веб-порталами.

Мы предлагаем контент-провайдерам разнообразные сервисы, созданные на базе хранилища семантических метаданных. Например:

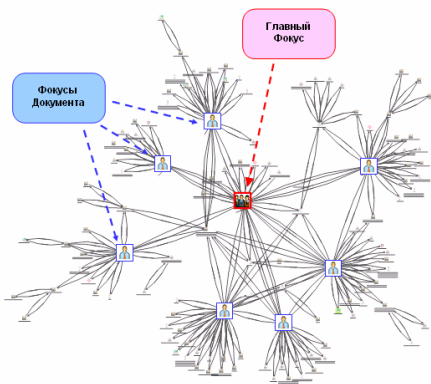
- сервис семантической разметки и навигации (см. выше)
- сервисы автоматического анализа информации:

- персонализированные RSS-дайджесты

возможность для пользователя под своим аккаунтом конструировать запросы, по которым доставляются новости — например, «высказывания об iPhone», «встречи/переговоры Дмитрия Медведева», «новости о ценных бумагах Газпрома»

- обобщенный отчет об объекте

автоматическое создание страницы со всей известной системе информацией о выбранном объекте (ссылки и выдержки из текстов по связям объекта, обнаруженным во всей коллекции обработанных ресурсов)



- **автоматический подбор связанного контента**

ссылки на новости на подобную тему (с аналогичными объектами и связями), ссылки на модели-конкуренты (в каталоге продуктов), позитивные/негативные высказывания о продуктах и т.д.

- **статистический анализ данных**

автоматическое построение графиков упоминаемости и значимости (по количеству связей) различных объектов с течением времени

Предлагаемые сервисы ориентированы на различные виды интернет-бизнеса.

**поиск**

генерация семантических метаданных для структуризации поисковой выдачи (см. выше), индексирование веб-ресурсов с помощью OntosMiner™

**интернет-магазины,  
каталоги**

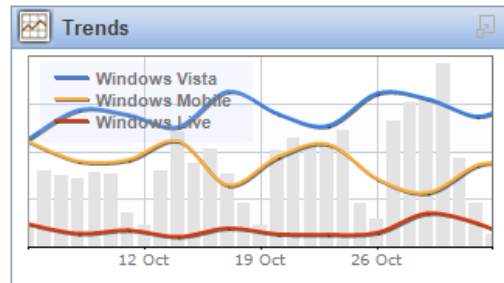
сервис автоматического подбора связанного контента (модели-конкуренты, поставщики, новости, высказывания о продукте и т.д.)

**информационные  
порталы**

сервис семантической разметки и навигации и автоматического анализа информации (см. выше)

Мы готовы к партнерству в разнообразных форматах — от полномасштабного проекта до бесплатного предоставления доступа к нашему API на тестовый период.

С наилучшими пожеланиями,  
команда AviComp Services



Российская компания Авикомп Сервисез (AviComp Services) входит в международный холдинг AviComp Group. Мы работаем на ИТ-рынке России, стран СНГ и Восточной Европы с 1992 года. Более пятнадцати лет мы занимаемся комплексной автоматизацией учреждений госсектора и крупных коммерческих компаний. Нашими партнерами являются мировые лидеры в области информационных технологий — Oracle, IBM, Hewlett Packard, SAP, Microsoft, 3Com, AVAYA Communication и др.

С 2001 года совместно со швейцарской компанией Ontos International мы развиваем технологии обработки естественного языка и предлагаем инновационные веб-решения в рамках концепции Семантического веба, пропагандируемой W3C.

ЗАО «Авикомп Сервисез»

просп. Вернадского 84, корп. 2  
119606 Москва

Т : (495) 436-09-45  
Ф : (495) 436-07-49  
[services@avicomp.com](mailto:services@avicomp.com)  
[www.avicomp.com](http://www.avicomp.com)

AviComp