

**ИЗВЛЕЧЕНИЕ СУЩНОСТЕЙ ИЗ ТЕКСТОВ В
СИСТЕМЕ КОНТЕНТ-МОНИТОРИНГА ИНТЕРНЕТ-
РЕСУРСОВ**

Ландэ Дмитрий Владимирович,
д.т.н., зам. директора EIVisti
Дармохвал Александр Теодорович,
аспирант, зав. отделом EIVisti

Партенит-2010

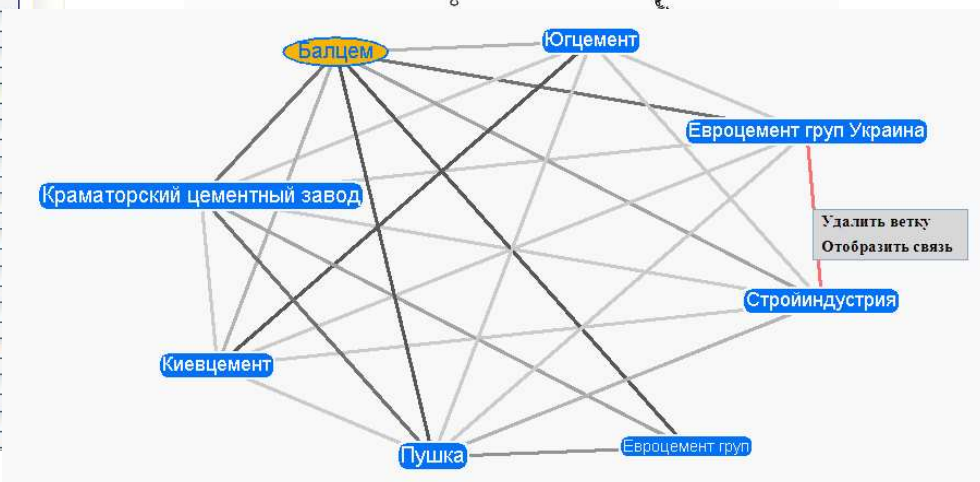
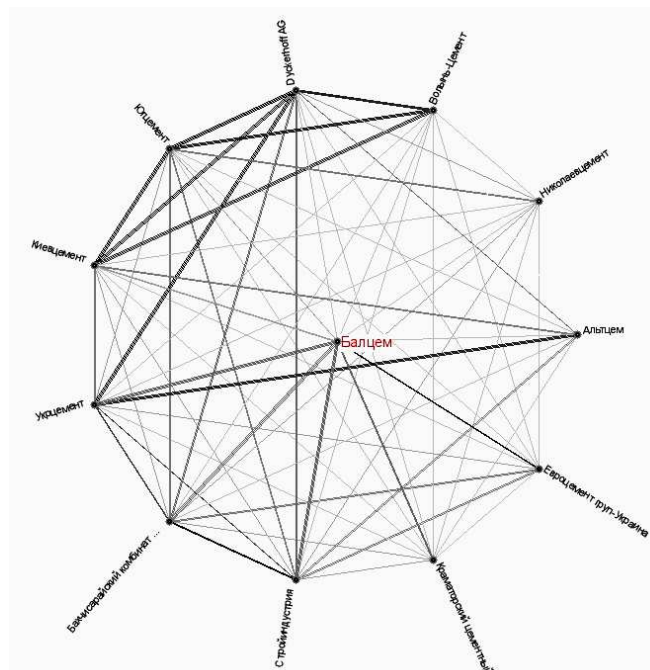
Использование сущностей, извлеченных из новостных сообщений

Информационный портрет
Уточнить запрос

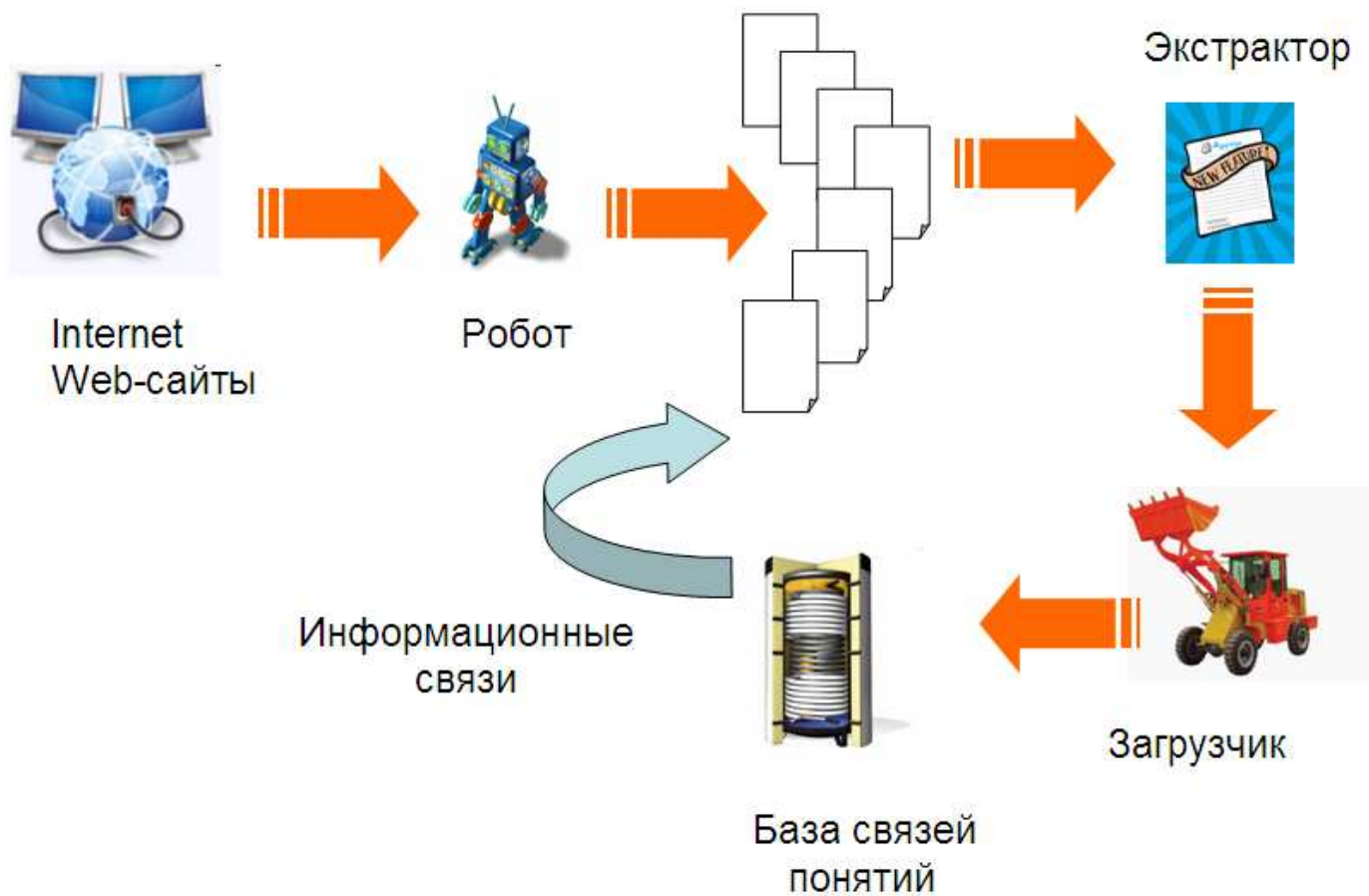
- Рубрики (36)
- Языки (3)
- Страны источников (19)
- Источники (50)
- Размер (3)
- Цифровая насыщенность (3)
- Тональность (2)
- География (50)
- Персоны (50)
- Компании (50)**

AND NOT

<input type="checkbox"/> BMW	<input type="checkbox"/>
<input type="checkbox"/> Volkswagen	<input type="checkbox"/>
<input type="checkbox"/> Ford	<input type="checkbox"/>
<input type="checkbox"/> ЕЭС России	<input type="checkbox"/>
<input type="checkbox"/> General Motors	<input type="checkbox"/>
<input type="checkbox"/> Opel	<input type="checkbox"/>
<input type="checkbox"/> Renault	<input type="checkbox"/>
<input type="checkbox"/> Интеко	<input type="checkbox"/>
<input type="checkbox"/> NBC Universal	<input type="checkbox"/>
<input type="checkbox"/> NBCU	<input type="checkbox"/>
<input type="checkbox"/> Verizon	<input type="checkbox"/>
<input type="checkbox"/> ИжАвто	<input type="checkbox"/>
<input type="checkbox"/> Toyota	<input type="checkbox"/>
<input type="checkbox"/> Крокус Интернейшнл	<input type="checkbox"/>
<input type="checkbox"/> Lamborghini	<input type="checkbox"/>



Технология экстрагирования сущностей в системе установления информационных связей



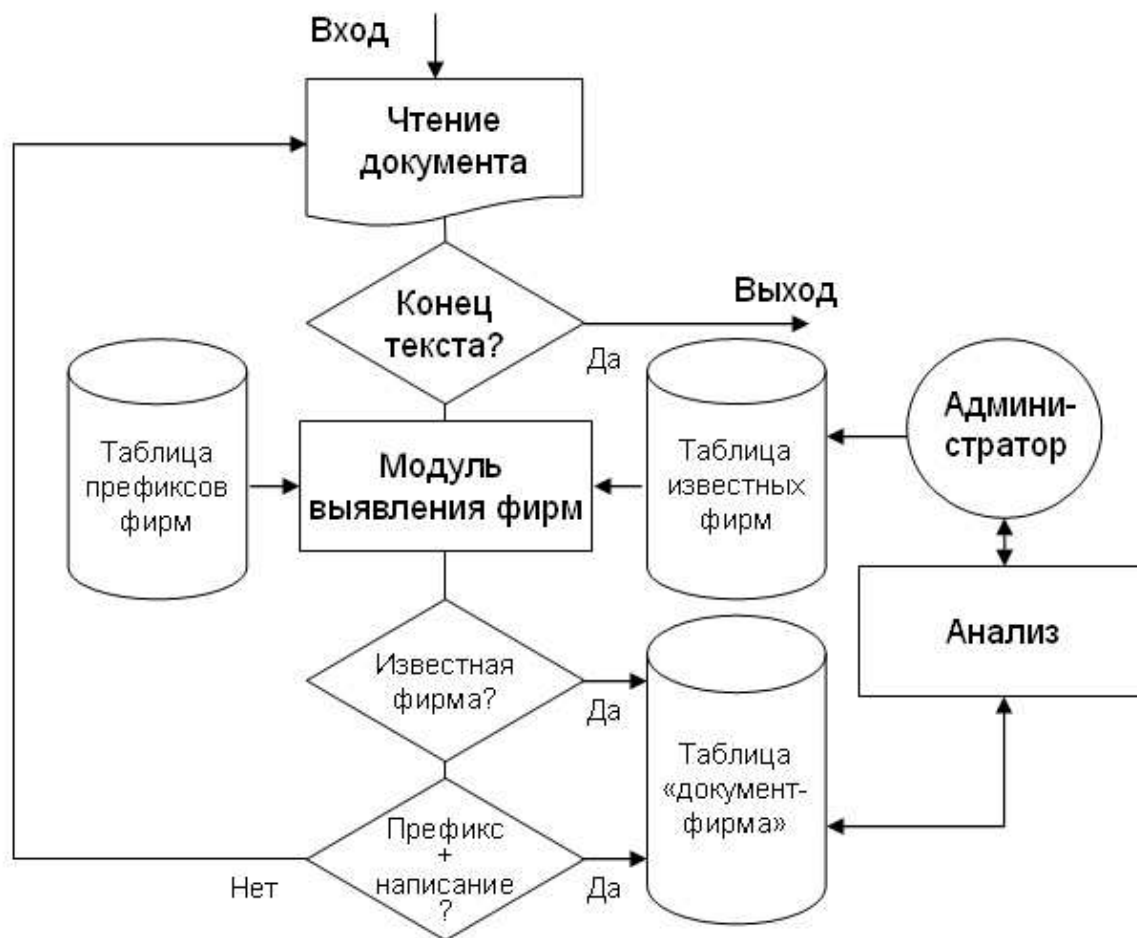
Применяемый метод экстрагирования сущностей –
«метод шаблонов и маркеров»

Маркеры: префиксы фирм: ТОВ, ООО, АО, АТ, АКБ...; должности: директор, акционер, председатель правления...; суффиксы: Ltd, Inc., Corp...

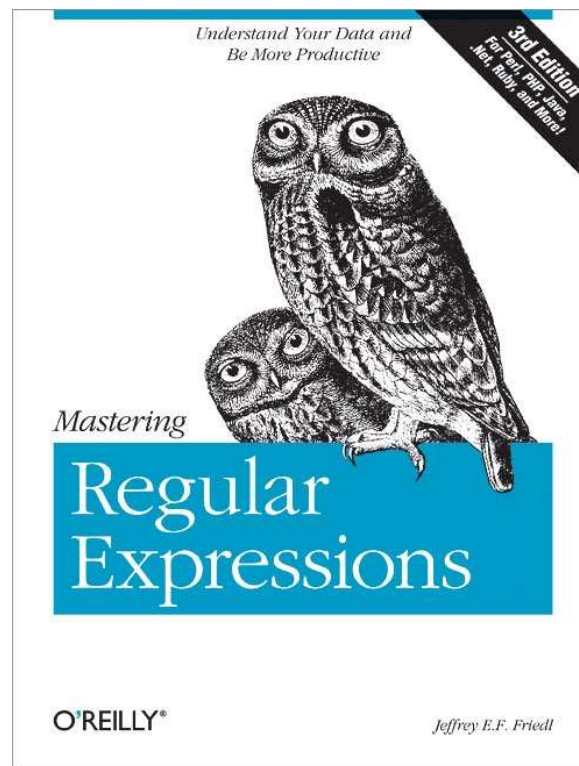
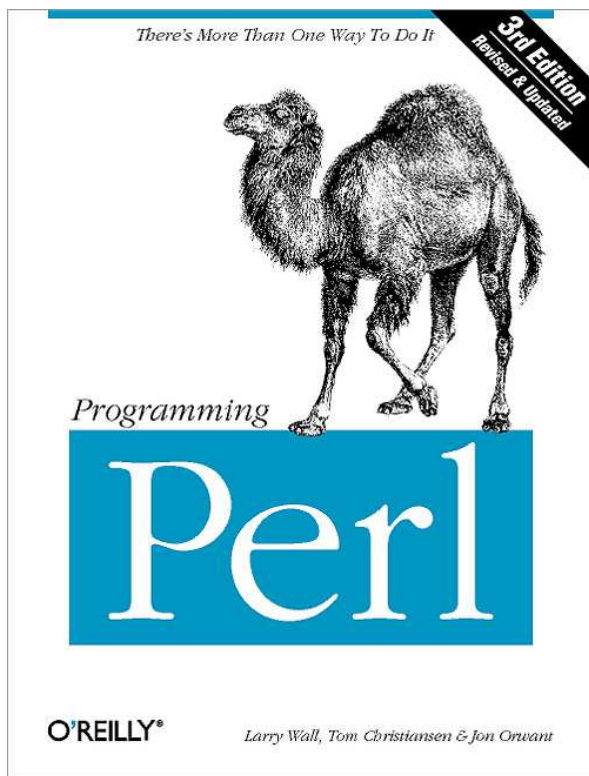
Шаблоны: правила написания сущностей; список заранее известных значений; список заведомо неправильных значений...

1. Индонезийская табачная компания **Djarum** первый поставщик сигарет с натуральной гвоздикой (кретек) в России.
2. Победителем аукциона признано **ООО "РГС Активы"**, предложившее за данный пакет акций начальную цену.
3. Он создан в конце 90-х годов и получил в свое управление по 50%+1 акция **ОАО "НПП "Каскад"** и **ОАО "АКБ "Південь"**, по 25%+1 акция **ОАО "Союз"** (Харьков) и **ОАО "Киевский проектный научно-опытный институт"**.
The Competition Commission got **Sasol Ltd.**, the world's biggest producer of motor fuel from coal, to sell almost all its fertilizer plants in July to boost competition.

Алгоритм выявления названий компаний



Программная реализация:
язык Perl и средства регулярных выражений RegExp.



Типы извлекаемых сущностей

Компании, Бренды

Персоны

Топонимы

Телефонные номера

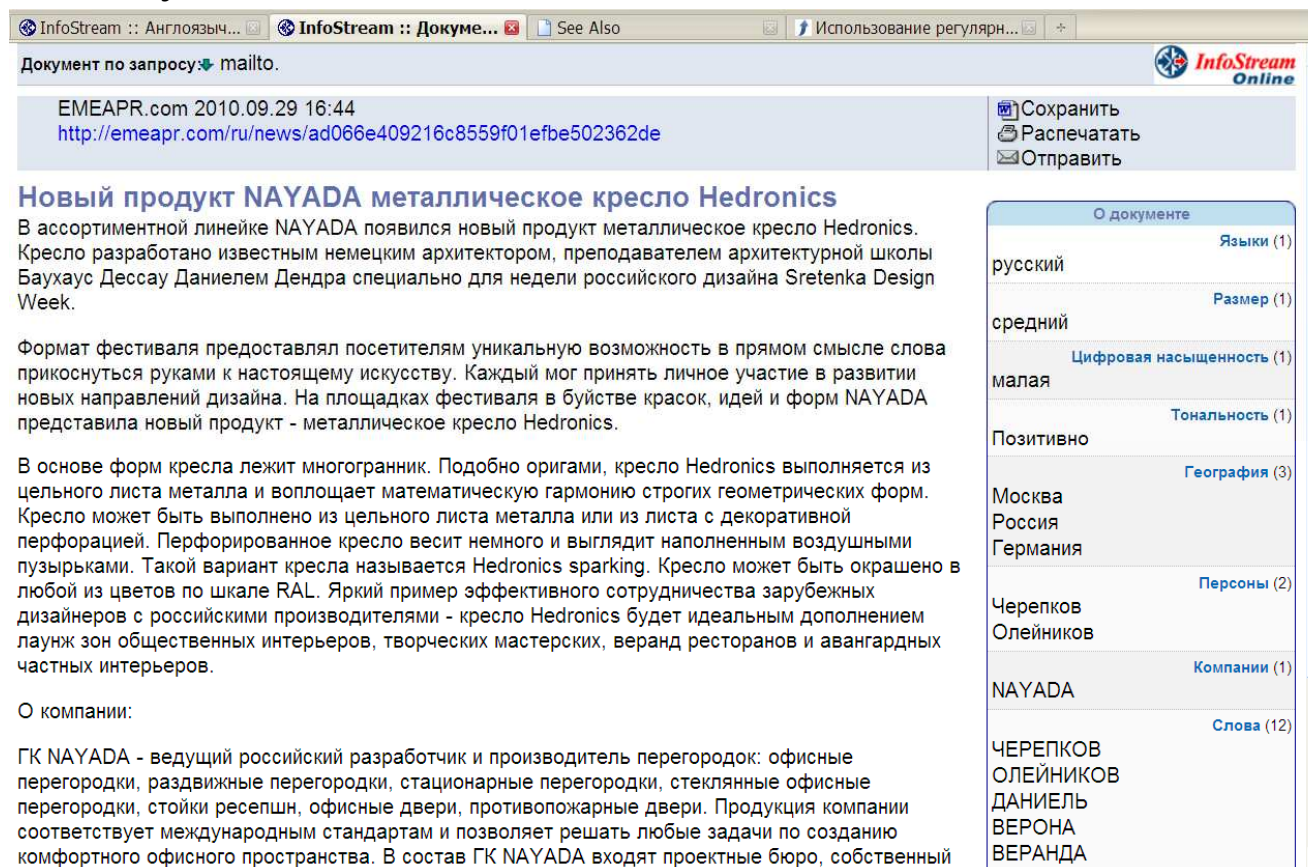
E-mail-адреса

WWW

Адреса

Коды ОКПО

Расчетные счета...



Документ по запросу: mailto.

EMEAPR.com 2010.09.29 16:44
<http://emeapr.com/ru/news/ad066e409216c8559f01efbe502362de>

Новый продукт NAYADA металлическое кресло Hedronics

В ассортиментной линейке NAYADA появился новый продукт металлическое кресло Hedronics. Кресло разработано известным немецким архитектором, преподавателем архитектурной школы Баухаус Дессау Даниелем Дендра специально для недели российского дизайна Sretenka Design Week.

Формат фестиваля предоставлял посетителям уникальную возможность в прямом смысле слова прикоснуться руками к настоящему искусству. Каждый мог принять личное участие в развитии новых направлений дизайна. На площадках фестиваля в буйстве красок, идей и форм NAYADA представила новый продукт - металлическое кресло Hedronics.

В основе форм кресла лежит многогранник. Подобно оригами, кресло Hedronics выполняется из цельного листа металла и воплощает математическую гармонию строгих геометрических форм. Кресло может быть выполнено из цельного листа металла или из листа с декоративной перфорацией. Перфорированное кресло весит немного и выглядит наполненным воздушными пузырьками. Такой вариант кресла называется Hedronics sparking. Кресло может быть окрашено в любой из цветов по шкале RAL. Яркий пример эффективного сотрудничества зарубежных дизайнеров с российскими производителями - кресло Hedronics будет идеальным дополнением лаунж зон общественных интерьеров, творческих мастерских, веранд ресторанов и авангардных частных интерьеров.

О компании:

ГК NAYADA - ведущий российский разработчик и производитель перегородок: офисные перегородки, раздвижные перегородки, стационарные перегородки, стеклянные офисные перегородки, стойки ресепшн, офисные двери, противопожарные двери. Продукция компании соответствует международным стандартам и позволяет решать любые задачи по созданию комфортного офисного пространства. В состав ГК NAYADA входят проектные бюро, собственный

Сохранить
Распечатать
Отправить

О документе

Языки (1)	русский
Размер (1)	средний
Цифровая насыщенность (1)	малая
Тональность (1)	Позитивно
География (3)	Москва Россия Германия
Персоны (2)	Черепков Олейников
Компании (1)	NAYADA
Слова (12)	ЧЕРЕПКОВ ОЛЕЙНИКОВ ДАНИЕЛЬ ВЕРОНА ВЕРАНДА

....Дополнительную информацию можно получить в пресс-службе NAYADA: (495) 933-10-55, 933-13-05, т.м.+7(906)076 8696 promo@nayada.ru Менеджер по связям с общественностью NAYADA - Людмила Олейникова г. Москва,

Бережковская набережная, д. 16а <http://www.nayada.ru/>

Основная проблема - быстродействие





*Горизонты прикладной лингвистики
и
лингвистических технологий*

СПАСИБО ЗА ВНИМАНИЕ!

Ландэ Дмитрий Владимирович,
dwl@visti.net

<http://infostream.ua>

<http://dwl.visti.net>

Киев-2010