ПЕРВЫЙ ОТКРЫТЫЙ КОНКУРС МОЛОДЫХ СПЕЦИАЛИСТОВ ЗАО «СИБКОТЭС»

Менеджмент, маркетинг и HR в энергетике

СИСТЕМА СИНТАКСИЧЕСКОГО ТЕМАТИЧЕСКОГО ПОИСКА В СЕТИ ИНТЕРНЕТ, СИСТЕМАТИЗАЦИИ И ПРЕДСТАВЛЕНИЯ КОНКУРСНЫХ ПРЕДЛОЖЕНИЙ И КОММЕРЧЕСКИХ НОВОСТЕЙ

Черноскутов Артем Сергеевич

E-mail: chernoskutov@inbox.ru

ЗАО "СибКОТЭС«

Новосибирск 2008

1. Использование сети Интернет для приска коммерческой информации

Что такое "коммерчески ценная информация"

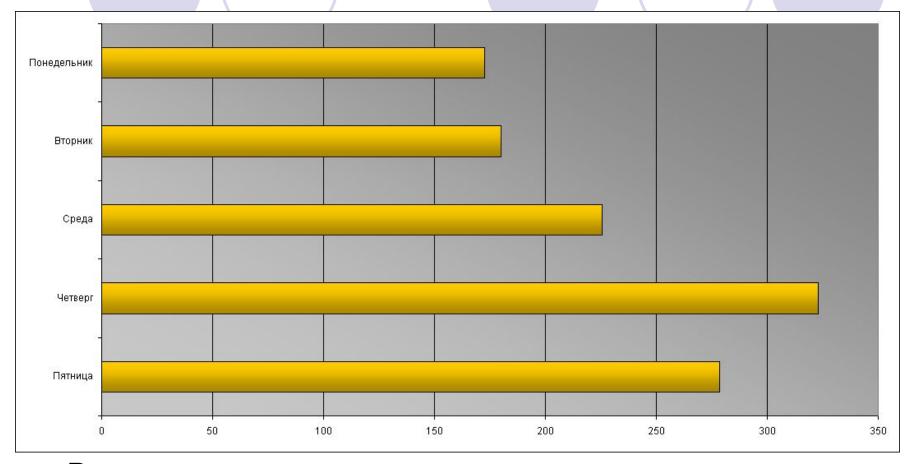
"горячие" конкурсы

Перспективные работы Перспективные Направления работ

3. Работа со списком конкурсов b2b-energo.ru

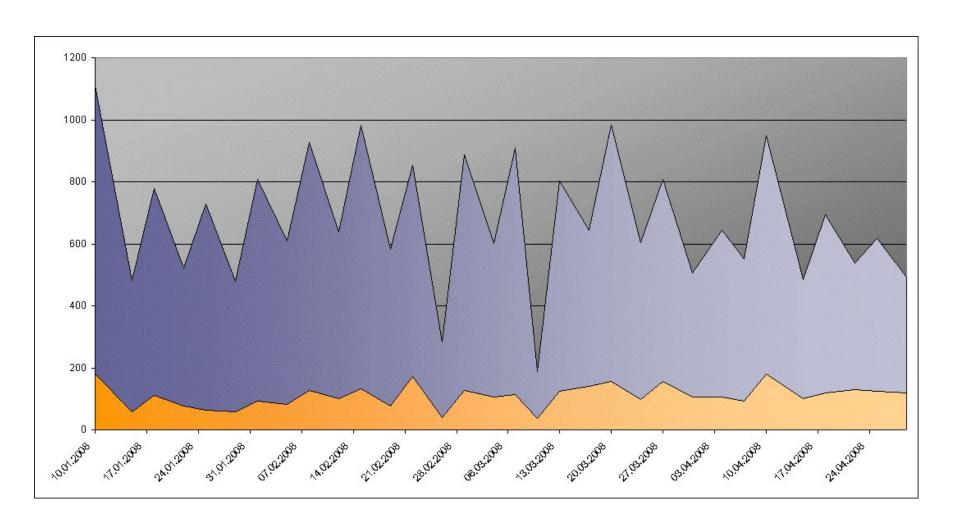
- http://www.b2b-energo.ru/summaries/copies.html
 B2B-Energo / Публикации о торгах / Копии публикаций
- http://www.b2b-energo.ru/summaries/summaries.html
 B2B-Energo / Публикации о торгах / Анонсирования торгов
- https://www.b2b-energo.ru/market/list.html?type=4
 B2B-Energo Торговая площадка Список объявлений о покупке
- https://www.b2b-energo.ru/market/list.html?type=2
 B2B-Energo Торговая площадка Список действующих аукционов покупателя
- https://www.b2b-energo.ru/market/list_tenders.html?open=1
 B2B-Energo Торговая площадка Список объявленных открытых конкурсов

Публикация данных

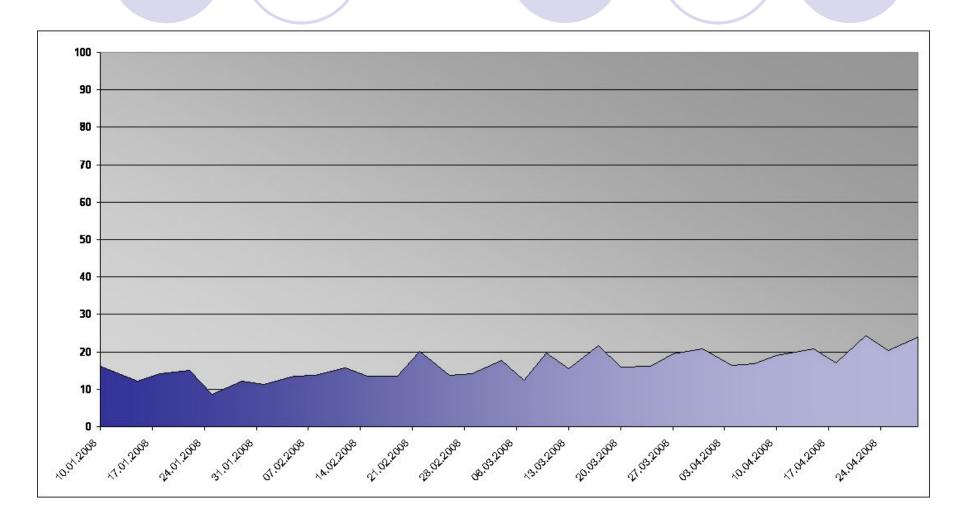


- Вечер понедельника утро вторника
- Вечер четверга утро пятницы

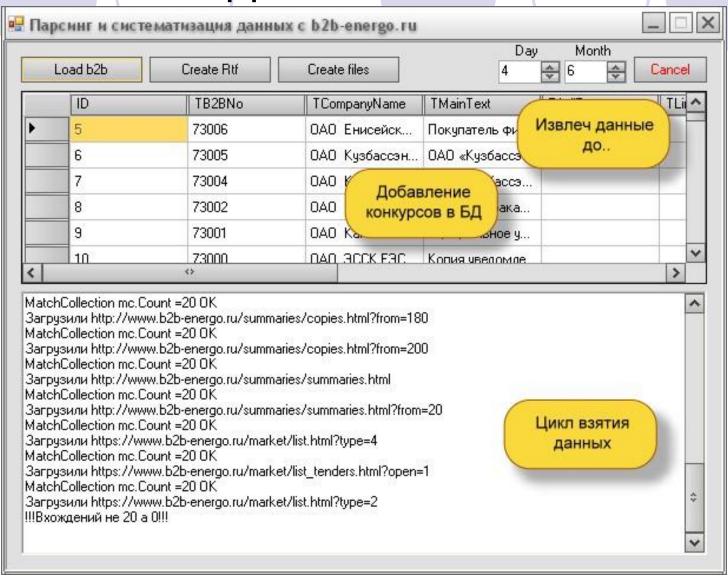
Сравнительный график общего числа опубликованных конкурсов (синий график) и числа конкурсов, включенных в отчет (оранжевый график).



Процентное соотношение между общим числом опубликованных конкурсов и числом конкурсов, взятых в отчет.



Цикл взятия данных



Возможные пути развития



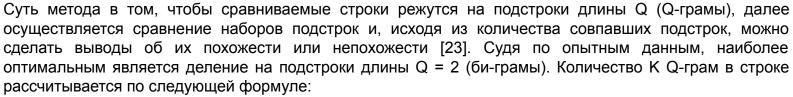
Одним из вариантов развития системы может быть оформление системы как web-сервиса и дополнение функциями контроля и управления. В частности, это может быть автоматическое взятие данных с b2b-enerdo по запросу, распределение конкурсов между пользователями, сопровождение конкурсов.

Сопровождение конкурсов может включать добавление комментариев или другую форму контроля, передачу конкурсов от одного пользователя другому, сбор статистической и прочей информации, оперативное изменение параметров поиска.

4. Работа с Новостями Электроэнергетики

- Поиск и обработка этих источников является масштабной работой поисковой системы
- В международной практике подборка опубликованных материалов, касающихся деятельности компании называется пресс - клиппинг (press-clipping, press-cutting, alert).

Алгоритм Q-грам.



К = Длина строки - Q + 1

Приведем небольшой пример реализации. Возьмем две строки:

"Строительство ТЭС" (эталон) и "ТЭС строится"

ст	тр	po	ОИ	ИТ	те	ел	ЛЬ	ьс	СТ	ТВ	во	0_	_T	тэ	эс
тэ	эс	c_	_c	ст	тр	po	ОИ	ИТ	тс	ся					

Совпадением считается одинаковый грам эталона и рабочей строки. Для примера они помечены цветом. Теперь определим критерий идентичности двумя способами:

КИ1 = Количество совпадений/ К эталона = 7/16 = 0.43

КИ2 = Количество совпадений * 2 / (К эталона + К рабочей строки) = 7*2/(16+11)= 0.52

Методика контекстно — зависимого поиска с применением технологии нейронных сетей

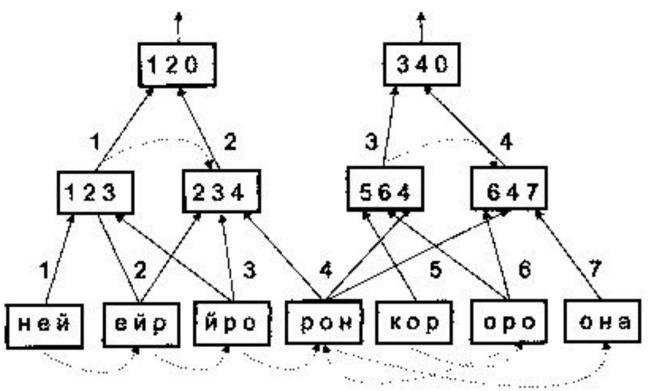


Рис. 3. Пример представления в трехуровневой нейронной сети слов "нейрон" и "корона" (рисунок взят из работы [21])



Спасибо!