

Участие ИК «Кодекс» в РОМИП 2004

Максим Губин

max@gubin.spb.ru

ИК «Кодекс»

Дорожки

- Поиск Web-коллекция
- Поиск нормативная коллекция

Прогоны

N	Коллекция	Гиперссылки	Взаимное положение слов
1	Web	-	-
2	Web	-	+
3	Legal	-	-
4	Legal	-	+
5	Legal	+	+

Учет взаимного положения слов

- Вес документа тем выше, чем выше вес входящих в него фрагментов, содержащих слова запроса.
- Вес каждого фрагмента тем выше, чем больше слов запроса он содержит.
- Чем ближе в фрагменте располагаются слова запроса, тем больший вес он имеет.
- Устойчивые словосочетания должны получать больший вес.

Отличия тестовой версии от коммерческой

- Не используются рубрикаторы и словари;
- Не используются шаблоны для разбора запроса.

Результаты

- Учет гипертекстовых ссылок улучшает качество;
- Учет взаимного положения – нет результата из-за ошибки реализации;
- Использование дополнительных атрибутов документа должно улучшить результаты.

Различие методик

РОМИП	Внутренняя
<ul style="list-style-type: none">• Большой объем обрабатываемой информации	<ul style="list-style-type: none">• Минимальный объем
<ul style="list-style-type: none">• Ориентация на TREC	<ul style="list-style-type: none">• Ориентация на usability
<ul style="list-style-type: none">• Система-Система	<ul style="list-style-type: none">• Человек-Система