

УЧАСТИЕ
ИНФОРМАЦИОННОЙ КОМПАНИИ
«КОДЕКС»
В РОМИП 2005

Максим Губин
max@gubin.spb.ru
«Кодекс»

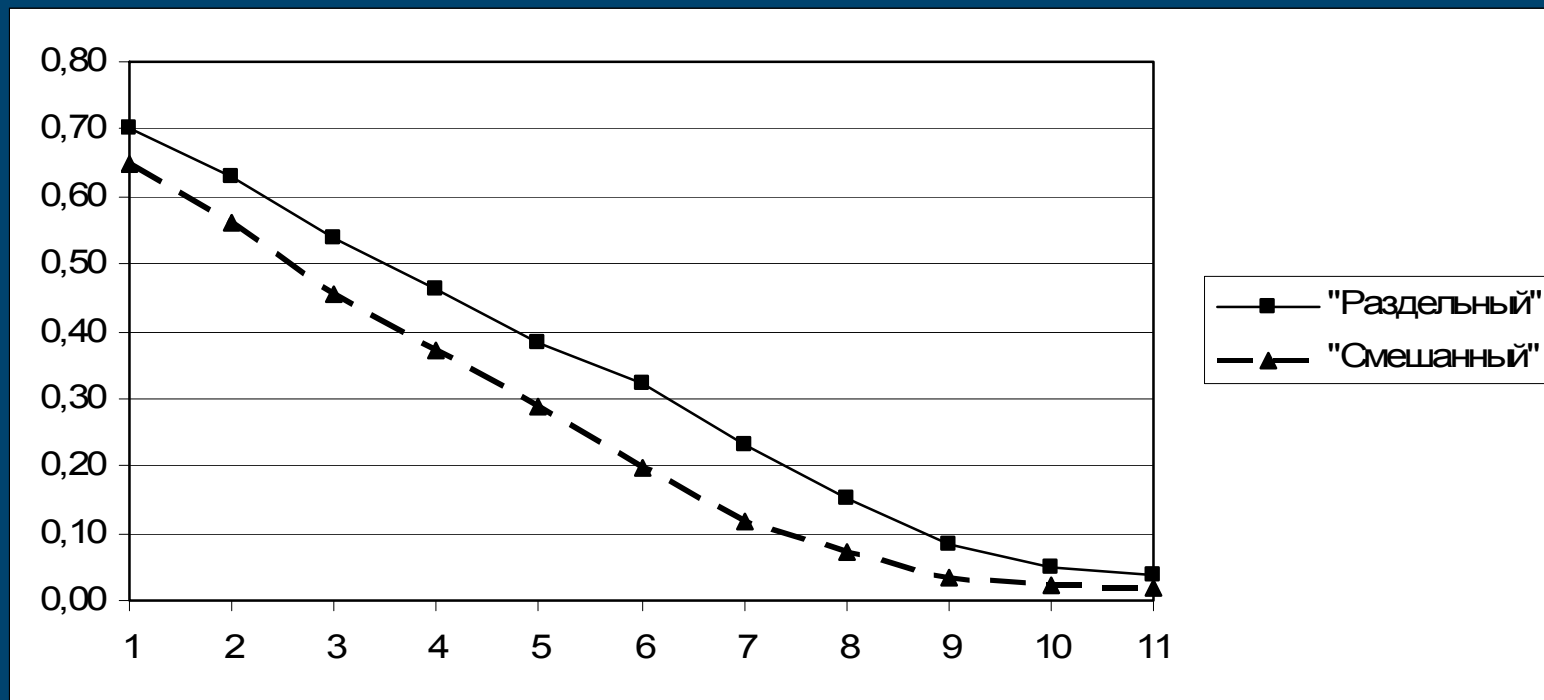
УЧАСТИЕ В ДОРОЖКАХ

- Web ad-hoc
- Legal ad-hoc
- Mixed ad-hoc
- Контекстно-зависимое аннотирование документов

MIXED AD-HOC

- «Сквозной» - все документы в одну коллекцию
- «Раздельный» - метапоиск по результатам каждой из коллекций

11-ТОЧЕЧНЫЙ ГРАФИК ТОЧНОСТЬ/ПОЛНОТА («СИЛЬНАЯ» РЕЛЕВАНТНОСТЬ)



КОНТЕКСТНО-ЗАВИСИМОЕ АННОТИРОВАНИЕ

- Прогон 1. Плотность слов запроса
- Прогон 2. Плотность слов запроса + наличие частотных слов документа
- Прогон 3. Плотность слов запроса + наличие слов, отобранных LRU-K алгоритмом

РЕЗУЛЬТАТЫ АННОТИРОВАНИЯ (AnnotationAccuracy)

	базовый	Freq	LRU-K
and and	0,789832	0,788395	0,788395
or or	0.893758	0.890610	0.890610

ВЫВОДЫ

1. Mixed ad-hoc «раздельный» оказался лучше
2. Контекстно-зависимое аннотирование –эксперимент не позволил выявить различий

ПРЕДЛОЖЕНИЯ РОМИП2006

- XML подобные форматы «обработанных» результатов
- Исходные тексты (готовые утилиты) для обсчета

**СПАСИБО
ЗА ВНИМАНИЕ**