

Russian Information Retrieval Evaluation Seminar (ROMIP)

<http://romip.ru/en/>

Igor Nekrestyanov, Pavel Braslavski

CLEF 2010

ROMIP at a glance

- TREC-like Russian initiative
- Started 2002
- Several text and image collections
- 10-15 participants per year (total 50+)
 - Academia and industry, students support
- ~3 000 man-hours of evaluation (2009)
- Remote participation + live meeting
- Collections are freely available
- Popular testbed for IR research in Russia
- Related activities: summer school in IR



Why?

- **Russia specifics**
 - Strong IR industry
 - Limited research in academia
 - Participation in global events considered complicated for Russian groups (language barrier, costs, etc.)
 - Russian language was not covered in international campaigns
- **Objectives**
 - Consolidate IR community
 - Stimulate research in the area
 - Independent evaluation

Evaluation methodology

- Similar to TREC approaches
- What's special?
 - Russian language collections
 - Some tasks are unique
 - E.g. news clustering, snippet generation, etc.
 - Mix of widely used and custom metrics
 - E.g. snippet informativeness/readability
 - Typically 2+ assessors (agreement 80-85%)
 - Domain experts for legal-related tracks
 - Rules and methodology are adjusted yearly

Largest text collections

Collection	Documents	Size (compressed)	Topics	Evaluated within ad-hoc search track
Legal	~300 000	2 Gb	14 794	220
By.Web	1 524 676	8 Gb	~ 60 000	1 500+
KM.RU	3 010 455	13 Gb	~ 60 000	~250

Text documents tracks

- Classic tracks run for years
 - Ad-hoc text retrieval
 - Text categorization (Web pages & sites, legal)
- Experimental tracks every year
 - Snippet generation
 - QA and fact extraction
 - News clustering
 - Search by sample document

Snippets evaluation

Romip Annotation Evaluation Tool

virtual dub К списку заданий

Цель - найти программу virtual dub. идеальный - ресурс с которого можно скачать программу virtual dub и описания к ней. релевантный+ - сайт с программами для монтажа видео. релевантный- - сайт с описаниями работы программ для монтажа видео; страницы с ссылками на сайты с программами; форумы с обсуждением работы программ для монтажа видео.

EveryDay - форум > обработка видео Запускаем Virtual Dub. Софт: а) программа Ulead Video Studio желательно 7-ой русский б) программа Virtual Dub 1.8 и выше в) установленные в системе кодеки, рекомендую пакет K-Lite Codec Pack 2.24 Full, Nimo Codec Pack, в инете их найти нет проблем г) необходима также и программа Quick Time Player дабы..	Документ релевантен? <input checked="" type="radio"/> Да <input type="radio"/> Нет <input type="radio"/> Не могу понять
EveryDay - форум > обработка видео 3. Софт: а) программа Ulead Video Studio желательно 7-ой русский б) программа Virtual Dub 1.8 и выше в) установленные в системе кодеки, рекомендую пакет K-Lite Codec Pack 2.24 Full, Nimo Codec Pack, в инете их найти нет проблем.	Ответ есть в заголовке документа? <input checked="" type="radio"/> Да <input type="radio"/> Нет <input type="radio"/> Не могу понять
	Полезна ли выделенная аннотация? Да: <input type="radio"/> Текст аннотации содержит ответ на вопрос <input type="radio"/> Понятно есть ли ответ в документе и текст хорошо читается <input checked="" type="radio"/> Можно догадаться есть ли ответ в документе, но текст сложно прочесть Нет: <input type="radio"/> Текст нормально читается, но из него непонятно содержит ли документ полезную информацию или нет <input type="radio"/> Текст аннотации состоит из кусочков, общий смысл непонятен <input type="radio"/> Текст невозможно прочесть

Image collections

- Photo collection: 20 000 images from **Flickr**
- Dups collection: 15 hrs video → 37 800 frames



21.09.2010



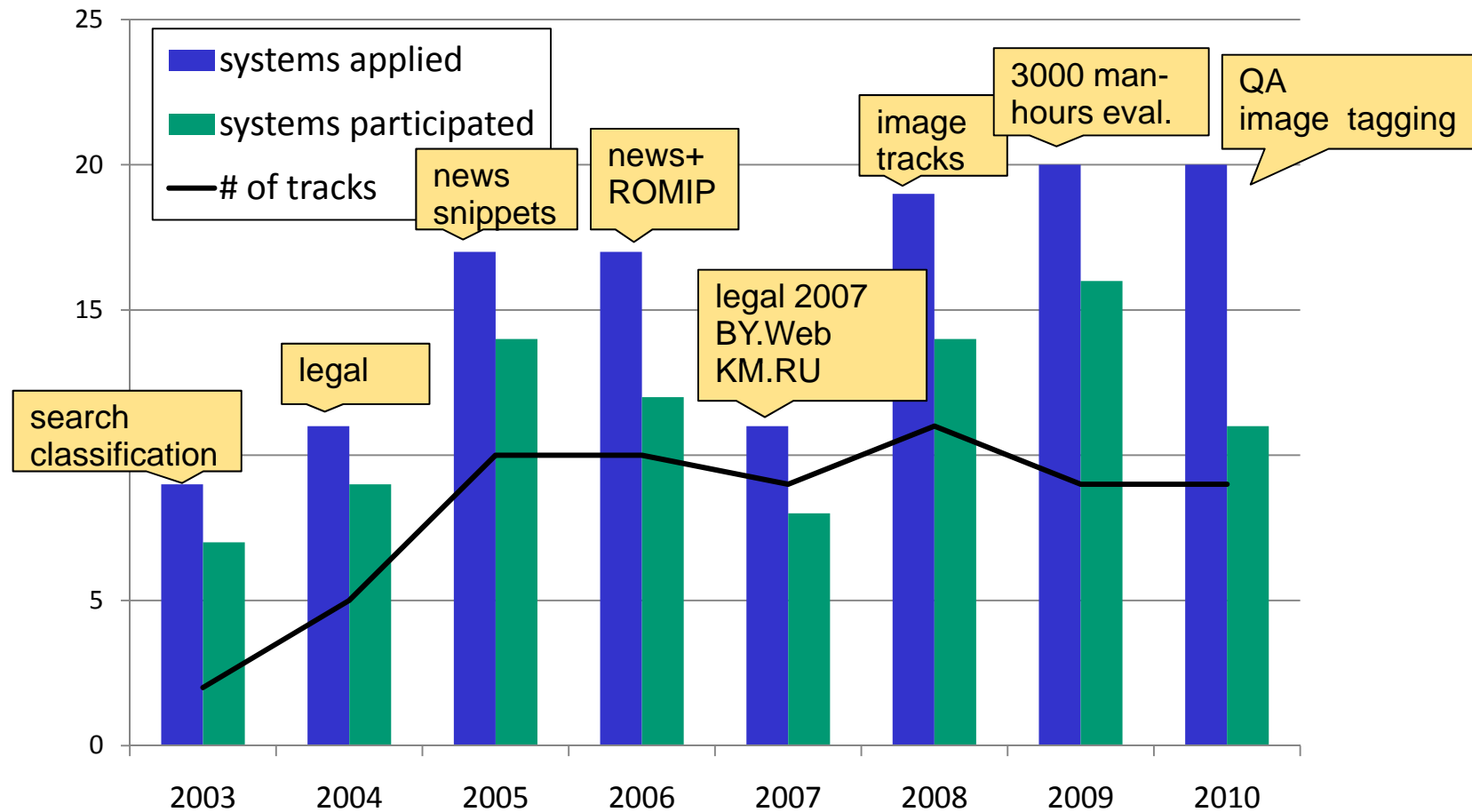
ROMIP

9

Image tracks

- Content based image retrieval (started 2008)
 - 750 tasks labeled
- Near-duplicate detection (started 2008)
 - ~1500 clusters
- Image annotation (started 2010)
 - ~ 1000 labeled images

ROMIP timeline



Thank you!

Questions?

Pavel Braslavski

pb@yandex-team.ru

Igor Nekrestyanov

romip@romip.ru

RuSSIR



21.09.2010

ROMIP

18