

# Experiments for the CL-SR task at CLEF 2006

**Muath Alzghool and Diana Inkpen**

University of Ottawa

Canada

Track: **Cross Language Spoken Retrieval (CL-SR)**



# Experiments

- Results for submitted runs - English collection
- Results for submitted runs - Czech collection
  - Segmentation issues, evaluation score
- Results for different systems: Smart, Terrier
- Query expansion
  - Log likelihood collocations scores
  - Terrier: divergence from randomness
  - Small improvements

# Results for the submitted runs for the English collection

Language	MAP	Fields	Description
English	0.2902	TDN	Terrier: MANUALKEYWORD + SUMMARY
English	0.0768	TDN	SMART: NSP query expansion (LL) ASRTEXT2004A + AUTOKEYWORD2004A1,A2
French	0.0637	TDN	SMART: ASRTEXT2004A + AUTOKEYWORD2004A1, A2
Spanish	0.0619	TDN	SMART: ASRTEXT2004A + AUTOKEYWORD2004A1, A2
English	<u>0.0565</u>	TD	Terrier: ASRTEXT2004A + ASRTEXT2006A + AUTOKEYWORD2004A1, A2

# Results for the submitted runs for the Czech collection

Language	GAP	Fields	Description
Czech	0.0039	TDN	SMART: ASRTEXT, CZECH AUTOKEYWORD, CZECH MANUKEYWORD, ENGLISH MANUKEYWORD, ENGLISH AUTOKEYWORD
Czech	0.0005	TDN	SMART: ASRTEXT, CZECHAUTOKEYWORD, CZECHEMANUKEYWORD
Czech	0.0004	TDN	SMART: ASRTEXT, CZECHAUTOKEYWORD
Czech	<u>0.0004</u>	TD	SMART: ASRTEXT, CZECHAUTOKEYWORD
Czech	<u>0.0005</u>	TD	Terrier: ASRTEXT, CZECHAUTOKEYWORD

# MAP scores for Terrier and SMART, with or without relevance feedback, for English topics

	System	Training			Test		
		TDN	TD	T	TDN	TD	T
1	SMART	<u>0.0954</u>	0.0906	0.0873	0.0766	0.0725	0.0759
	SMARTnsp	0.0923	0.0901	0.0870	<u>0.0768</u>	<u>0.0754</u>	<u>0.0769</u>
2	Terrier	0.0913	0.0834	0.0760	0.0651	0.0560	0.0656
	TerrierKL	0.0915	<u>0.0952</u>	<u>0.0906</u>	0.0654	0.0565	0.0685

# Experiments

- Various ASR transcripts (2003, 2004, 2006)
  - New ASR 2006 transcripts do not help
  - Combinations do not help
  - Automatic keywords help
- Cross-language
  - Results good for French to English topic translations
  - Not for Spanish, German, Czech
- Manual summaries and manual keywords
  - Best results

# MAP scores for **Terrier**, with various ASR transcript combinations

Segment fields	Training			Test		
	TDN	TD	T	TDN	TD	T
ASR2003A	0.0733	0.0658	0.0684	0.0560	0.0473	0.0526
ASR 2004A	0.0794	0.0742	0.0722	<u>0.0670</u>	0.0569	0.0604
ASR 2006A	0.0799	0.0731	0.0741	0.0656	0.0575	0.0576
ASR 2006B	0.0840	0.0770	0.0776	0.0665	<u>0.0576</u>	0.0591
ASR 2003A+2004A	0.0759	0.0722	0.0705	0.0596	0.0472	0.0542
ASR 2004A+2006A	0.0811	0.0743	0.0730	0.0638	0.0492	0.0559
ASR 2004A+2006B	0.0804	0.0735	0.0732	0.0628	0.0494	0.0558
ASR 2003A+AUTOK	0.0873	0.0859	0.0789	0.0657	0.0570	0.0671
ASR 2004A+AUTOK	0.0915	<u>0.0952</u>	0.0906	0.0654	0.0565	0.0685
ASR 2006B+AUTOK	<u>0.0926</u>	0.0932	0.0909	0.0717	0.0608	0.0661
ASR 2004A+2006A+AUTOK	0.0915	<u>0.0952</u>	<u>0.0925</u>	0.0654	0.0565	<u>0.0715</u>
ASR 2004A+2006B+AUTOK	0.0899	0.0909	0.0890	0.0640	0.0556	0.0692

# MAP scores for **Smart**, with various ASR transcript combinations

Segment fields	Training			Test		
	TDN	TD	T	TDN	TD	T
ASR 2003A	0.0625	0.0586	0.0585	0.0508	0.0418	0.0457
ASR 2004A	0.0701	0.0657	0.0637	0.0614	0.0546	0.0540
ASR 2006A	0.0537	0.0594	0.0608	0.0455	0.0434	0.0491
ASR 2006B	0.0582	0.0635	0.0642	0.0484	0.0459	0.0505
ASR 2003A+2004A	0.0685	0.0646	0.0636	0.0533	0.0442	0.0503
ASR 2004A+2006A	0.0686	0.0699	0.0696	0.0543	0.0490	0.0555
ASR 2004A+2006B	0.0686	0.0713	0.0702	0.0542	0.0494	0.0553
ASR 2003A +AUTOK	0.0923	0.0847	0.0839	0.0674	0.0616	0.0690
ASR 2004A+AUTOK	<u>0.0954</u>	0.0906	0.0873	<u>0.0766</u>	<u>0.0725</u>	<u>0.0759</u>
ASR 2006B+AUTOK	0.0869	0.0892	0.0895	0.0650	0.0659	0.0734
ASR2004A+2006A+AUTOK	0.0903	<u>0.0932</u>	0.0915	0.0654	0.0654	0.0777
ASR2004A+2006B+AUTOK	0.0895	0.0931	<u>0.0919</u>	0.0652	0.0655	0.0742



# Results of the cross-language experiments

	Language	Training			Test		
		TDN	TD	T	TDN	TD	T
1	<b>English</b>	<b><u>0.0954</u></b>	<b><u>0.0906</u></b>	<b><u>0.0873</u></b>	<b><u>0.0766</u></b>	<b><u>0.0725</u></b>	<b><u>0.0759</u></b>
2	<b>French</b>	<b>0.0950</b>	<b>0.0904</b>	<b>0.0814</b>	<b>0.0637</b>	<b>0.0566</b>	<b>0.0483</b>
3	<b>Spanish</b>	<b>0.0773</b>	<b>0.0702</b>	<b>0.0656</b>	<b>0.0619</b>	<b>0.0589</b>	<b>0.0488</b>
4	<b>German</b>	<b>0.0653</b>	<b>0.0622</b>	<b>0.0611</b>	<b>0.0674</b>	<b>0.0605</b>	<b>0.0618</b>
5	<b>Czech</b>	<b>0.0585</b>	<b>0.0506</b>	<b>0.0421</b>	<b>0.0400</b>	<b>0.0309</b>	<b>0.0385</b>

- Indexed fields ASRTEXT2004, and autokeywords
- using SMART with the weighting scheme Inn.ntn

# Results of indexing the **manual keywords and summaries**, using SMART with weighting scheme Inn.ntn, and Terrier with In(exp)C2

	Language / System	Training			Test		
		TDN	TD	T	TDN	TD	T
1	English SMART	0.3097	0.2829	0.2564	0.2654	0.2344	0.2258
2	English Terrier	<u>0.3242</u>	<u>0.3227</u>	<u>0.2944</u>	<u>0.2902</u>	<u>0.2710</u>	<u>0.2489</u>
3	French SMART	0.2920	0.2731	0.2465	0.1861	0.1582	0.1495
4	French Terrier	0.3043	0.3066	0.2896	0.1977	0.1909	0.1651
5	Spanish SMART	0.2502	0.2324	0.2108	0.2204	0.1779	0.1513
6	Spanish Terrier	0.2899	0.2711	0.2834	0.2444	0.2165	0.1740
7	German SMART	0.2232	0.2182	0.1831	0.2059	0.1811	0.1868
8	German Terrier	0.2356	0.2317	0.2055	0.2294	0.2116	0.2179
9	Czech SMART	0.1766	0.1687	0.1416	0.1275	0.1014	0.1177
10	Czech Terrier	0.1822	0.1765	0.1480	0.1411	0.1092	0.1201

# Conclusion and future work

- ✦ Low retrieval results, except when using manual summaries and keywords
- ✦ Future work
  - Filter out potential speech errors – semantic outliers with low PMI score (in a large Web corpus) with neighboring words
  - Index using speech lattices