ecir 2018 40th
European Conference on Information Retrieval
Monday 26th - Thursday 29th March 2018 Grenoble, France

Statistical Stemmers: A Reproducibility Study

Gianmaria Silvello, Riccardo Bucco, Giulio Busato, Giacomo Fornari, Andrea Langeli, Alberto Purpura, Giacomo Rocco, Alessandro Tezza, and Maristella Agosti

Department of Information Engineering,
University of Padua, Italy
{name.surname}@unipd.it

ECIR 2018

@giansilv

Outline

Statistical stemmers

Goal and organisation of the work

The reproducibility study

Conclusions

Statistical Stemmers

Statistical Stemmers

- Language-independent (good for low-resources languages)

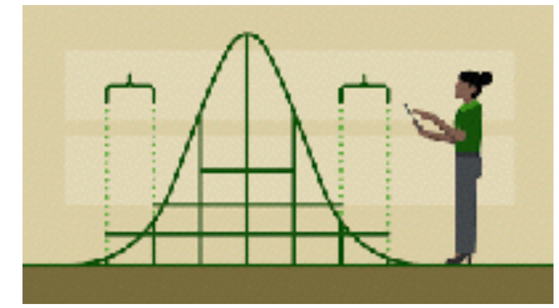
- Not readily available in off-the-shelf IR systems

- Not usually taken into account in longitudinal studies in IR

- New stemmers proposed in the last 5-10 years

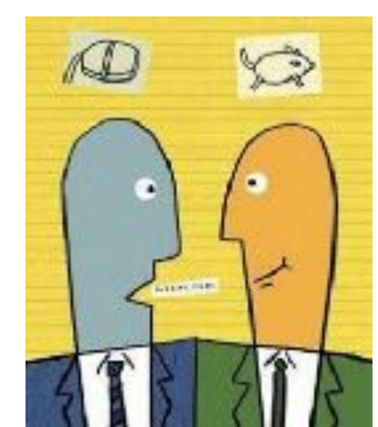
Goal of the work

- Realise statistical stemmers



- Ready to use stemmers in an off-the-shelf IR system (Terrier IR system)

- Focus on core stemmers aggregated by time (2011) and context (authors and target languages)



Focus

- Paik, J.H., Parui, S.K.: (**FCB**) A Fast Corpus-Based Stemmer. ACM Trans. Asian Lang. Inf. Process. 10(2), 1-16 (2011)
- Paik, J.H., Pal, D., Parui, S.K.: (**SNS**) A Novel Corpus-based Stemming Algorithm Using Co-occurrence Statistics. In: SIGIR 2011. pp. 863-872. ACM Press (2011)
- Paik, J.H., Mitra, M., Parui, S.K., Jarvelin, K. **GRAS**: An effective and efficient stemming algorithm for information retrieval. ACM Trans. Inf. Syst. 29(4), 19 (2011)

Goal and Organization

Reproduce to Learn

- 2017-2018 master course in Information Retrieval @ University of Padua

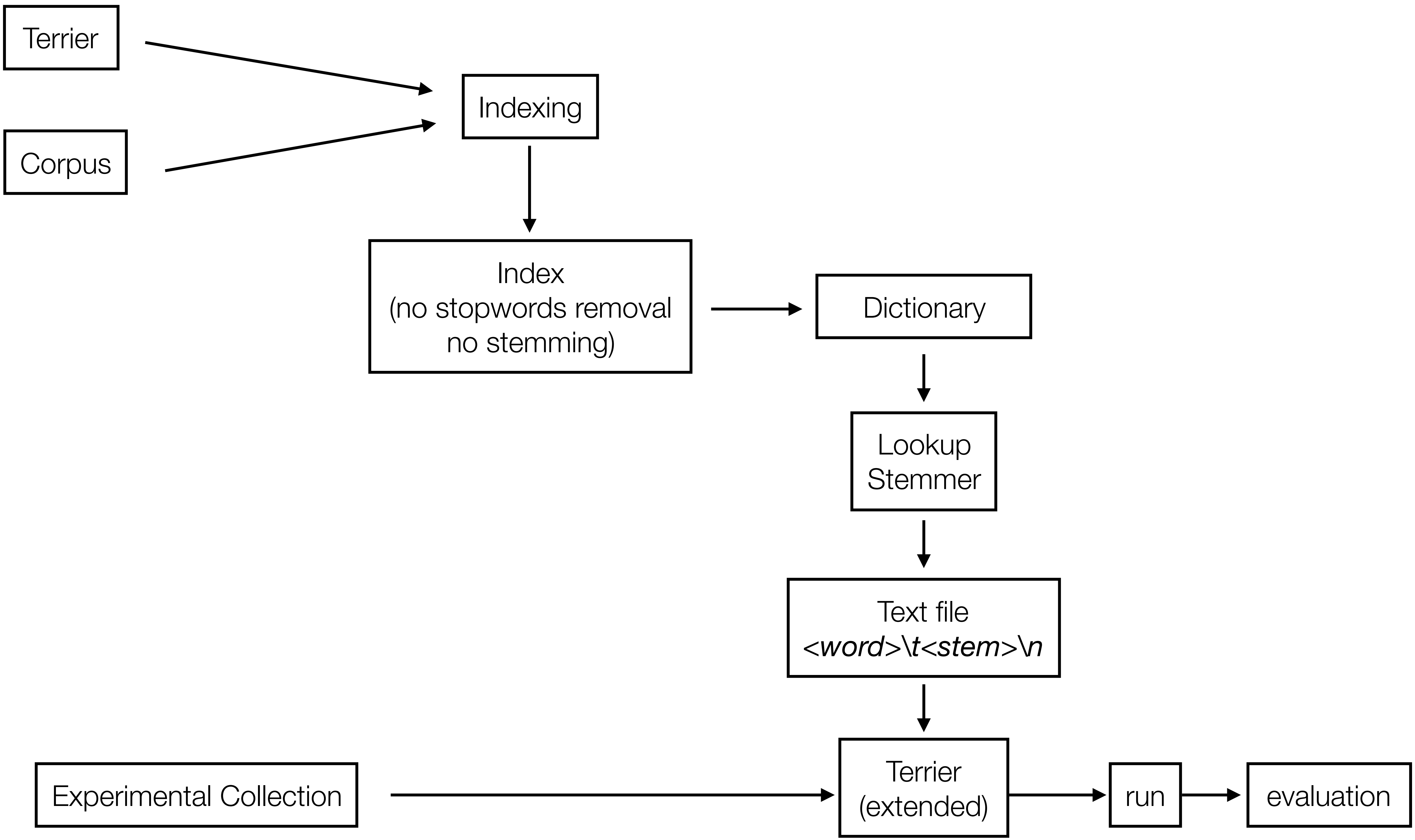
- 16 groups of students (1-4 people each)

- 10 statistical stemmers to be realised

Statistical Stemmers

- **FCB**: Paik, J. H. and Parui, S. K. (2011). A Fast Corpus-Based Stemmer. ACM Trans. Asian Lang. Inf. Process., 10(2):8.
- **GRAS**: Paik, J. H., Mitra, M., Parui, S. K., and Jarvelin, K. (2011). GRAS: An effective and efficient stemming algorithm for information retrieval. ACM Trans. Inf. Syst., 29(4):19.
- HPS: Brychcin, T. and Konopik, M. (2015). HPS: High precision stemmer. Inf. Process. Manage., 51(1):68–91.
- OARD: Oard, D. W., Levow, G., and Cabezas, C. I. (2000). CLEF Experiments at Maryland: Statistical Stemming and Backoff Translation. In Peters, C., editor, Cross-Language Information Retrieval and Evaluation: Workshop of Cross-Language Evaluation Forum (CLEF 2000), pages 176– 187. Lecture Notes in Computer Science (LNCS) 2069, Springer, Heidelberg, Germany.
- **SNS**: Paik, J. H., Pal, D., and Parui, S. K. (2011). A Novel Corpus- based Stemming Algorithm Using Co-occurrence Statistics. In Ma, W.-Y., Nie, J.-Y., Baeza-Yaetes, R., Chua, T.-S., and Croft, W. B., editors, Proc. 34th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2011), pages 863–872. ACM Press, New York, USA.
- SPLIT: Bacchin, M., Ferro, N., and Melucci, M. (2005). A Probabilistic Model for Stemmer Generation. Information Processing & Management, 41(1):121–137.
- STON: Melucci, M. and Orio, N. (2003). A Novel Method for Stemmer Generation Based on Hidden Markov Models. In Kraft, D., Frieder, O., Hammer, J., Qureshi, S., and Seligman, L., editors, Proc. 12th International Conference on Information and Knowledge Management (CIKM 2003), pages 131–138. ACM Press, New York, USA.
- YASS: Majumder, P., Mitra, M., Parui, S. K., Kole, G., Mitra, P., and Datta, K. (2007). YASS: Yet Another Suffix Stripper. ACM Transactions on Information Systems (TOIS), 25(4):18:1–3:20.
- XU: Xu, J. and Croft, W. B. (1998). Corpus-Based Stemming Using Cooccurrence of Word Variants. ACM Trans. Inf. Syst., 16(1): 61–81.

Steps of the work



https://github.com/giansilv/statisticalStemmers

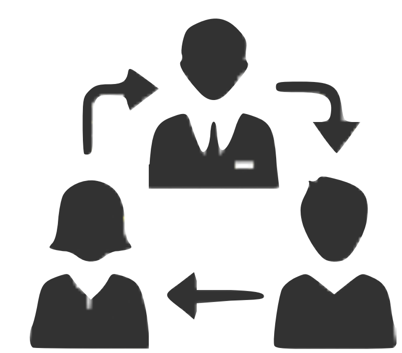
Steps of the work

- Work for the course (semi-supervised learning)

  - Choose a stemmer, implement the stemmer, evaluate it on a shared test collection (i.e. CLEF-IT-2003)

- Work for the reproducibility study (supervised learning)

  1. Select best groups and restrict to volunteers



  2. Re-implement the stemmers

  3. Evaluate the stemmers on different test collections, compare results and, possibly, go back to step 2

The Reproducibility Study

FCB: How it works

- Identify all suffixes and define sets of words sharing the same suffix → set with cardinality **α** is called potential suffix

- Defines k-equivalence classes of words with the same prefix → iterative process starting with length 5 and going down to 2

- Evaluation of the prefix as stem for each class

    - determine the size of the subsets of elements in the class that contains only terms whose suffixes all belong to the potential suffixes set

    - determine the ratio between these sets and the potential classes

    - if ratio > **δ** then the longest prefix is a stem for all the terms in the class

FCB: Implementation details

- No description of the suffixes extraction process:
  - we extracted all suffixes without considering the inclusion relations between them
- Underspecified choice: A potential-class is defined as the largest subset of words with a common prefix R ending with frequent suffixes:
  - FCB v.1: considers the strings composed of all the characters that follow the common prefix R as suffixes of the terms in a class
  - FCB v.2: considers the whole terms in a k -equivalence class and qualifies them as ending with a frequent suffix if they end with any of the frequent suffixes mentioned before (we allow chars between prefix and suffix)

FCB: Implementation details

- No description of the suffixes extraction process:
  - we extracted all suffixes without considering the inclusion relations between them

No description about how to deal with singleton classes; singleton classes have as longest common prefix the whole term that belongs to the class, therefore, the induced suffix is always empty. **This implies that the terms with a unique prefix, but ending with a frequent suffix are not stemmed.**

  - FCB v.1: considers the strings composed of all the characters that follow the common prefix R as suffixes of the terms in a class

  - FCB v.2: considers the whole terms in a k -equivalence class and qualifies them as ending with a frequent suffix if they end with any of the frequent suffixes mentioned before (we allow chars between prefix and suffix)

FCB: Reproducibility

- Experimental collections:
  - Hungarian: CLEF 2006-2007 with 98 topics
  - English: TIPSTER Wall Street Journal sub-corpus (topics 1-200)
- Topics: two sets, 1 only T and 1 T+D
- Retrieval model: IFB2
- We tested FCB v.1 and v.2 on several values of δ ratio and we selected the closest performing one

FCB: Hungarian

- Closest results with δ=0.5 and FCB v.2

		Original			Reproduced			Difference		
		MAP	RPrec	P@10	MAP	RPrec	P@10	MAP	RPrec	P@10
T	No Stem	0.185	0.199	0.258	0.1830	0.1956	0.2547	-0.0020	-0.0034	-0.0033
	FCB	0.293	0.315	0.353	0.2863	0.2942	0.3284	-0.0067	**-0.0208**	**-0.0246**
	RB	0.267	0.280	0.343	0.2610	0.2737	0.3245	-0.0060	-0.0063	**-0.0185**
TD	No Stem	0.239	0.252	0.314	0.2375	0.2528	0.3133	-0.0015	+0.0008	-0.0007
	FCB	0.341	0.352	0.390	0.3355	0.3263	0.3949	-0.0055	**-0.0257**	+0.0049
	RB	0.335	0.340	0.389	0.3347	0.3358	0.4102	-0.0020	**-0.0532**	**+0.0212**

FCB: Hungarian

- Closest results with δ=0.5 and FCB v.2

		Original			Reproduced			Difference		
		MAP	RPrec	P@10	MAP	RPrec	P@10	MAP	RPrec	P@10
T	No Stem	0.185	0.199	0.258	0.1830	0.1956	0.2547	-0.0020	-0.0034	-0.0033
	FCB	0.293	0.315	0.353	0.2863	0.2942	0.3284	-0.0067	**-0.0208**	**-0.0246**
	RB	0.267	0.280	0.343	0.2610	0.2737	0.3245	-0.0060	-0.0063	**-0.0185**
TD	No Stem	0.239	0.252	0.314	0.2375	0.2528	0.3133	-0.0015	+0.0008	-0.0007
	FCB	0.341	0.352	0.390	0.3355	0.3263	0.3949	-0.0055	**-0.0257**	+0.0049
	RB	0.335	0.340	0.389	0.3347	0.3358	0.4102	-0.0020	**-0.0532**	**+0.0212**

1. There are minor differences also when no stem or a “standard” rule-based stemmer is applied (differences due to Terrier versions?)

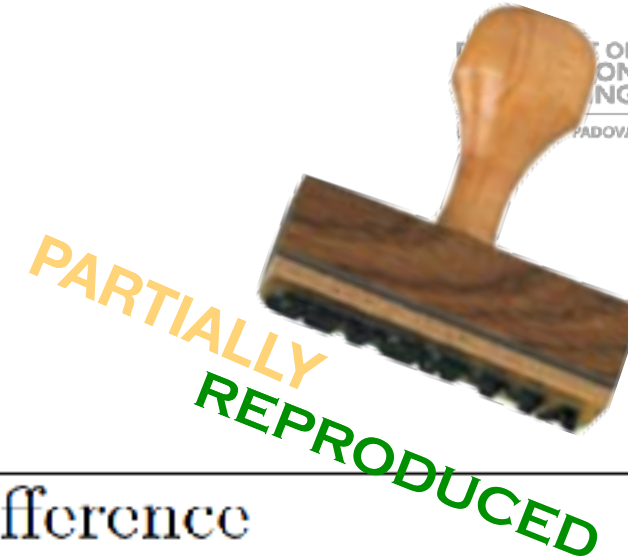
FCB: Hungarian

- Closest results with δ=0.5 and FCB v.2

		Original			Reproduced			Difference		
		MAP	RPrec	P@10	MAP	RPrec	P@10	MAP	RPrec	P@10
T	No Stem	0.185	0.199	0.258	0.1830	0.1956	0.2547	-0.0020	-0.0034	-0.0033
	FCB	0.293	0.315	0.353	0.2863	0.2942	0.3284	-0.0067	**-0.0208**	**-0.0246**
	RB	0.267	0.280	0.343	0.2610	0.2737	0.3245	-0.0060	-0.0063	**-0.0185**
TD	No Stem	0.239	0.252	0.314	0.2375	0.2528	0.3133	-0.0015	+0.0008	-0.0007
	FCB	0.341	0.352	0.390	0.3355	0.3263	0.3949	-0.0055	**-0.0257**	+0.0049
	RB	0.335	0.340	0.389	0.3347	0.3358	0.4102	-0.0020	**-0.0532**	**+0.0212**

2. Performance differences are consistent across different stemming approaches

FCB: English



- Closest results with δ=0.6 and FCB v.1

		Original			Reproduced			Difference		
		MAP	RPrec	P@10	MAP	RPrec	P@10	MAP	RPrec	P@10
T	No Stem	0.225	0.267	0.399	0.2250	0.2674	0.3990	0.000	0.000	0.000
	FCB	0.258	0.289	0.437	0.2399	0.2791	0.4020	**-0.018**	**-0.010**	**-0.035**
	Porter	0.261	0.296	0.432	0.2621	0.2971	0.4362	+0.001	+0.001	+0.004
TD	No Stem	0.272	0.312	0.477	0.2722	0.3125	0.4765	0.000	+0.000	0.000
	FCB	0.295	0.331	0.493	0.2811	0.3181	0.4715	**-0.014**	**-0.013**	**-0.0215**
	Porter	0.294	0.325	0.477	0.2958	0.3262	0.4800	+0.002	+0.001	+0.0030

Performance differences are smaller than for the Hungarian collection but FCB version is not consistent

No stemmer and Porter stemmer approaches lead to almost perfect results → why this does not happen for Hungarian?

SNS: How it works

- Computation of the co-occurrence strength of word pairs
    - A weighted graph is built where two words are connected if they have a common prefix of length l1
- Re-calculation of the strengths
    - The strength assigned to a word pair (w1;w2) is proportional to the number of other words in the corpus that co-occur with both w1 and w2
- Clustering of the words

SNS: Implementation details

- Step 1. Extract the data from the lexicon and the inverted index and discard digits and terms shorter than <u>l1 = 3</u>
- Step 2. Compute the co-occurrence between two words if they have a common prefix with length greater than or equal to l1; if their strength is not zero, we check if their common prefix is greater than or equal to <u>l2 = 5</u>
- Step 3. Create a weighted graph where words are nodes and the edges are weighted by their co-occurrence strength
- Step 4. Update the edge strength by re-calculating the co-occurrence of terms
- Step 5. Remove the non-strong edges (clustering phase)
- Step 6. Find the connected components of the graph

SNS: Implementation details

Each stem is generated by finding the longest common prefix amongst the connected words; this is an **assumption** we made since this phase is not described and the proposed algorithm stops right after the creation of word clusters.

SNS: Reproducibility

- Experimental collections:
    - CLEF Bulgarian 2006-2007 (100 topics)
    - CLEF Hungarian 2006-2007 (98 topics)
    - CLEF Czech 2007 (49 topics)
    - TREC TIPSTER Disk 4&5 minus CR (150 topics)
- Topics: one set T+D
- Retrieval model: IFB2

SNS: Czech

	Original			Reproduced			Difference		
	NO	RB	SNS	NO	RB	SNS	NO	RB	SNS
MAP	0.2381	0.3409	0.3624	0.2382	0.3405	0.3569	+0.0001	-0.0004	-0.0055
RPrec	0.2611	0.3456	0.3441	0.2611	0.3456	0.3449	0.0000	0.0000	-0.0008
P@10	0.2680	0.3480	0.3700	0.2680	0.3480	0.3640	0.0000	0.0000	-0.0060

The authors claim to use the stemmer defined in [Savoy, 2009], which actually presents two rule-based stemmers, one light and one aggressive.

We tested both these stemmers and we found that the light one was used in the reference paper.

SNS: Czech

	Original			Reproduced			Difference		
	NO	RB	SNS	NO	RB	SNS	NO	RB	SNS
MAP	0.2381	0.3409	0.3624	0.2382	0.3405	0.3569	+0.0001	-0.0004	-0.0055
RPrec	0.2611	0.3456	0.3441	0.2611	0.3456	0.3449	0.0000	0.0000	-0.0008
P@10	0.2680	0.3480	0.3700	0.2680	0.3480	0.3640	0.0000	0.0000	-0.0060

The authors claim to use the stemmer defined in [Savoy, 2009], which actually presents two rule-based stemmers, one light and one aggressive.

We tested both these stemmers and we found that the light one was used in the reference paper.

SNS: Bulgarian

	Original			Reproduced			Difference		
	NO	RB	SNS	NO	RB	SNS	NO	RB	SNS
MAP	0.2166	0.2794	0.3256	0.2038	0.2786	0.2980	**-0.0128**	-0.0008	**-0.0276**
RPrec	0.2293	0.2930	0.3289	0.2291	0.3033	0.3253	-0.0002	**+0.0103**	-0.0036
P@10	0.2570	0.3270	0.3520	0.2580	0.3410	0.3540	+0.0010	**+0.0140**	+0.0020

SNS: Bulgarian

	Original			Reproduced			Difference		
	NO	RB	SNS	NO	RB	SNS	NO	RB	SNS
MAP	0.2166	0.2794	0.3256	0.2038	0.2786	0.2980	**-0.0128**	-0.0008	**-0.0276**
RPrec	0.2293	0.2930	0.3289	0.2291	0.3033	0.3253	-0.0002	**+0.0103**	-0.0036
P@10	0.2570	0.3270	0.3520	0.2580	0.3410	0.3540	+0.0010	**+0.0140**	+0.0020

Rule-based stemmer not specified. 3 choices:

1) aggressive stemmer
2) light stemmer using transliterated terms
3) light stemmer processing documents in Cyrillic

closest performances to the reference paper

SNS: Bulgarian

	Original			Reproduced			Difference		
	NO	RB	SNS	NO	RB	SNS	NO	RB	SNS
MAP	0.2166	0.2794	0.3256	0.2038	0.2786	0.2980	**-0.0128**	-0.0008	**-0.0276**
RPrec	0.2293	0.2930	0.3289	0.2291	0.3033	0.3253	-0.0002	**+0.0103**	-0.0036
P@10	0.2570	0.3270	0.3520	0.2580	0.3410	0.3540	+0.0010	**+0.0140**	+0.0020

SNS is successfully reproduced, but there are sizeable differences for the rule-based stemmer and also for no stemmer

SNS: Bulgarian

	Original			Reproduced			Difference		
	NO	RB	SNS	NO	RB	SNS	NO	RB	SNS
MAP	0.2166	0.2794	0.3256	0.2038	0.2786	0.2980	**-0.0128**	-0.0008	**-0.0276**
RPrec	0.2293	0.2930	0.3289	0.2291	0.3033	0.3253	-0.0002	**+0.0103**	-0.0036
P@10	0.2570	0.3270	0.3520	0.2580	0.3410	0.3540	+0.0010	**+0.0140**	+0.0020

SNS is successfully reproduced, but there are sizeable differences for the rule-based stemmer and also for no stemmer

SNS improves the baseline systems both in the reference paper and in the reproduced version, even though in the reproduced case the improvement is less marked

SNS: Bulgarian

	Original			Reproduced			Difference		
	NO	RB	SNS	NO	RB	SNS	NO	RB	SNS
MAP	0.2166	0.2794	0.3256	0.2038	0.2786	0.2980	**-0.0128**	-0.0008	**-0.0276**
RPrec	0.2293	0.2930	0.3289	0.2291	0.3033	0.3253	-0.0002	**+0.0103**	-0.0036
P@10	0.2570	0.3270	0.3520	0.2580	0.3410	0.3540	+0.0010	**+0.0140**	+0.0020

SNS is successfully reproduced, but there are sizeable differences for the rule-based stemmer and also for no stemmer

SNS improves the baseline systems both in the reference paper and in the reproduced version, even though in the reproduced case the improvement is less marked

SNS: Hungarian

	Original			Reproduced			Difference		
	NO	RB	SNS	NO	RB	SNS	NO	RB	SNS
MAP	0.2386	0.3132	0.3588	0.2375	0.3369	0.3583	-0.0011	**+0.0237**	-0.0005
RPrec	0.2518	0.3117	0.3585	0.2528	0.3459	0.3556	+0.0010	**+0.0342**	-0.0029
P@10	0.3143	0.3990	0.4224	0.3133	0.4153	0.4163	-0.0010	**+0.0163**	-0.0061

The authors specified a rule-based stemmer [Savoy, 2008] where a light and an aggressive stemmers are defined. The closest results are obtained with the light one.

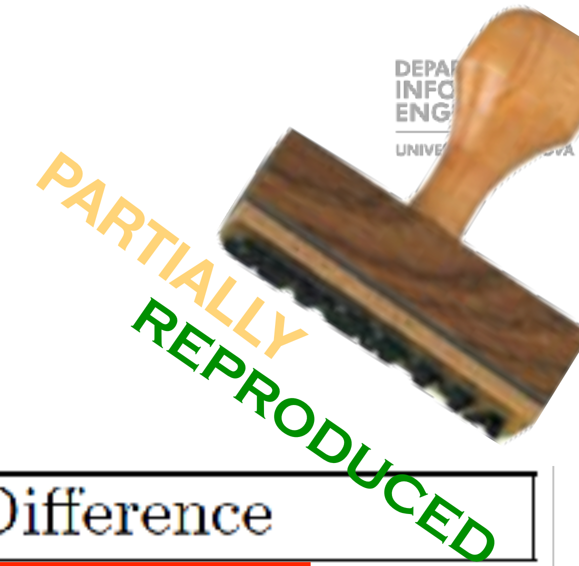
SNS: Hungarian

	Original			Reproduced			Difference		
	NO	RB	SNS	NO	RB	SNS	NO	RB	SNS
MAP	0.2386	0.3132	0.3588	0.2375	0.3369	0.3583	-0.0011	**+0.0237**	-0.0005
RPrec	0.2518	0.3117	0.3585	0.2528	0.3459	0.3556	+0.0010	**+0.0342**	-0.0029
P@10	0.3143	0.3990	0.4224	0.3133	0.4153	0.4163	-0.0010	**+0.0163**	-0.0061

The authors specified a rule-based stemmer [Savoy, 2008] where a light and an aggressive stemmers are defined. The closest results are obtained with the light one.

SNS is still slightly superior to the rule-based stemmer we employed even though this **difference is less marked** than the one reported in the reference paper

SNS: Hungarian



	Original			Reproduced			Difference		
	NO	RB	SNS	NO	RB	SNS	NO	RB	SNS
MAP	0.2386	0.3132	0.3588	0.2375	0.3369	0.3583	-0.0011	**+0.0237**	-0.0005
RPrec	0.2518	0.3117	0.3585	0.2528	0.3459	0.3556	+0.0010	**+0.0342**	-0.0029
P@10	0.3143	0.3990	0.4224	0.3133	0.4153	0.4163	-0.0010	**+0.0163**	-0.0061

The authors specified a rule-based stemmer [Savoy, 2008] where a light and an aggressive stemmers are defined. The closest results are obtained with the light one.

SNS is still slightly superior to the rule-based stemmer we employed even though this **difference is less marked** than the one reported in the reference paper

SNS: English

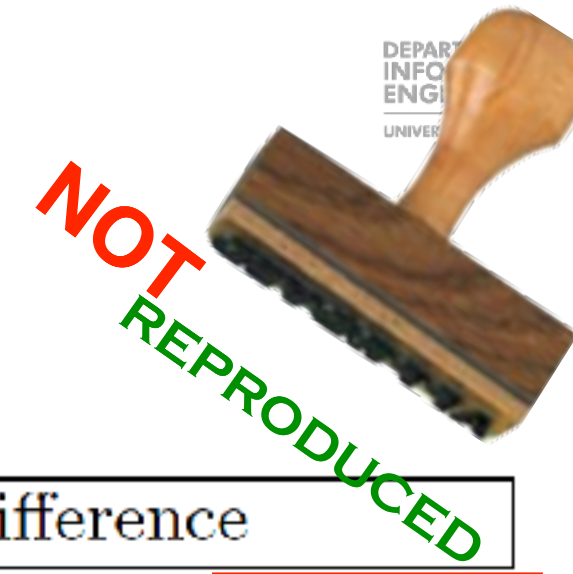
	Original			Reproduced			Difference		
	NO	RB	SNS	NO	RB	SNS	NO	RB	SNS
MAP	0.2290	0.2599	0.2582	0.2289	0.2596	0.2319	-0.0001	-0.0003	**-0.0263**
RPrec	0.2733	0.3008	0.3001	0.2736	0.3013	0.2722	+0.0003	+0.0005	**-0.0279**
P@10	0.4327	0.4833	0.4727	0.4320	0.4827	0.4267	-0.0007	-0.0006	**-0.0460**

SNS: English

	Original			Reproduced			Difference		
	NO	RB	SNS	NO	RB	SNS	NO	RB	SNS
MAP	0.2290	0.2599	0.2582	0.2289	0.2596	0.2319	-0.0001	-0.0003	**-0.0263**
RPrec	0.2733	0.3008	0.3001	0.2736	0.3013	0.2722	+0.0003	+0.0005	**-0.0279**
P@10	0.4327	0.4833	0.4727	0.4320	0.4827	0.4267	-0.0007	-0.0006	**-0.0460**

SNS for English is not successfully reproduced and its performances are lower than in the reference paper

SNS: English



	Original			Reproduced			Difference		
	NO	RB	SNS	NO	RB	SNS	NO	RB	SNS
MAP	0.2290	0.2599	0.2582	0.2289	0.2596	0.2319	-0.0001	-0.0003	**-0.0263**
RPrec	0.2733	0.3008	0.3001	0.2736	0.3013	0.2722	+0.0003	+0.0005	**-0.0279**
P@10	0.4327	0.4833	0.4727	0.4320	0.4827	0.4267	-0.0007	-0.0006	**-0.0460**

SNS for English is not successfully reproduced and its performances are lower than in the reference paper

GRAS: How it works

- Identify the word partitions sharing a **l**-long prefix; l is set to be the average word length for the given language

- Determine the common suffixes of the words sharing a prefix

  - Two suffixes are considered a candidate pair if they are shared "frequently enough" by word pairs: **α parameter**

- Create a graph where the identified words are mapped to nodes which are connected by an edge if the words are morphologically related

- Create equivalence classes on the basis of a cohesion **parameter δ**

GRAS: Implementation details

- The **l** parameter is set to be “the average word length for the language concerned”, but no further details are given

- We chose to calculate the average length of the words in the lexicon

GRAS: Implementation details

- The l parameter is set to be "the average word length for the language concerned", but no further details are given

- We chose to calculate the average length of the words in the lexicon

	Original					Reproduced				
	EN	FR	HU	BG	CZ	EN	FR	HU	BG	CZ
Docs	472,525	177,452	49,530	87,281	81,735	472,525	177,452	49,530	69,281	81,735
Words	522,381	303,349	528,315	320,673	457,164	502,280	325,292	534,813	292,077	457,149
						-3.84%	-7.23%	-1.23%	-9%	-0.01%

Nicola Ferro, Gianmaria Silvello: 3.5K runs, 5K topics, 3M assessments and 70M measures: What trends in 10 years of Adhoc-ish CLEF? Inf. Process. Manage. 53(1): 175-202 (2017)

GRAS: Implementation details

- The **l** parameter is set to be “the average word length for the language concerned”, but no further details are given

- We chose to calculate the average length of the words in the lexicon

	Original					Reproduced				
	EN	FR	HU	BG	CZ	EN	FR	HU	BG	CZ
Docs	472,525	177,452	49,530	87,281	81,735	472,525	177,452	49,530	69,281	81,735
Words	522,381	303,349	528,315	320,673	457,164	502,280	325,292	534,813	292,077	457,149

-3.84% -7.23% -1.23% -9% -0.01%

	Czech	Bulgarian	English	French	Hungarian
l	6	7	7	7	8

Nicola Ferro, Gianmaria Silvello: 3.5K runs, 5K topics, 3M assessments and 70M measures: What trends in 10 years of Adhoc-ish CLEF? Inf. Process. Manage. 53(1): 175-202 (2017)

GRAS: Reproducibility

- Experimental collections:
  - CLEF Bulgarian 2006-2007 (100 topics)
  - CLEF French 2005-2006 (100 topics)
  - CLEF Hungarian 2006-2007 (98 topics)
  - CLEF Czech 2007 (49 topics)
  - TREC TIPSTER Disk 4&5 minus (CR+FR) (150 topics)
- Topics: one set T+D
- Retrieval model: IFB2

GRAS: Reproducibility

		MAP	R-Prec	P@5	P@10	Rel Ret
BG	Original	0.3260	0.3340	0.4240	0.3550	2110
	Reproduced	0.3410	0.3580	0.4730	0.3720	2043
	Diff	**+0.0150**	**+0.0240**	**+0.0490**	**+0.0170**	−67
CZ	Original	0.3660	0.3600	0.4480	0.3760	689
	Reproduced	0.3630	0.3580	0.4460	0.3720	690
	Diff	−0.0030	−0.0020	−0.0020	−0.0040	+1
EN	Original	0.2700	0.3090	0.5430	0.4790	7873
	Reproduced	0.2749	0.3128	0.5492	0.4859	7904
	Diff	+0.0049	+0.0038	+0.0062	+0.0069	+31
FR	Original	0.3870	0.3980	0.5330	0.4910	4078
	Reproduced	0.3867	0.3886	0.5495	0.4838	4115
	Diff	−0.0003	−0.0094	**+0.0165**	−0.0072	+37
HU	Original	0.3510	0.3600	0.4740	0.4220	1924
	Reproduced	0.3319	0.3467	0.4701	0.4104	1846
	Diff	**-0.0191**	**-0.0133**	−0.0039	**-0.0116**	−78

BG: unique terms (-9%)
HIGH DIFFERENCE
HIGH impact

FR: unique terms (-7.23%)
HIGH DIFFERENCE
LOW IMPACT

HU: unique terms (-1.23%)
LOW DIFFERENCE
HIGH impact

GRAS: Reproducibility

		MAP	R-Prec	P@5	P@10	Rel Ret
BG	Original	0.3260	0.3340	0.4240	0.3550	2110
	Reproduced	0.3410	0.3580	0.4730	0.3720	2043
	Diff	**+0.0150**	**+0.0240**	**+0.0490**	**+0.0170**	−67
CZ	Original	0.3660	0.3600	0.4480	0.3760	689
	Reproduced	0.3630	0.3580	0.4460	0.3720	690
	Diff	−0.0030	−0.0020	−0.0020	−0.0040	+1
EN	Original	0.2700	0.3090	0.5430	0.4790	7873
	Reproduced	0.2749	0.3128	0.5492	0.4859	7904
	Diff	+0.0049	+0.0038	+0.0062	+0.0069	+31
FR	Original	0.3870	0.3980	0.5330	0.4910	4078
	Reproduced	0.3867	0.3886	0.5495	0.4838	4115
	Diff	−0.0003	−0.0094	**+0.0165**	−0.0072	+37
HU	Original	0.3510	0.3600	0.4740	0.4220	1924
	Reproduced	0.3319	0.3467	0.4701	0.4104	1846
	Diff	**-0.0191**	**-0.0133**	−0.0039	**-0.0116**	−78

unique terms (-9%)
HIGH DIFFERENCE
HIGH impact

unique terms (-7.23%)
HIGH DIFFERENCE
LOW IMPACT

unique terms (-1.23%)
LOW DIFFERENCE
HIGH impact

Conclusions

Conclusions

- Employment of a standard open-source system as Terrier for the experiments

Conclusions

- Employment of a standard open-source system as Terrier for the experiments
    - System version not reported

Conclusions

- Employment of a standard open-source system as Terrier for the experiments
  - System version not reported
- Use standard test collections

Conclusions

- Employment of a standard open-source system as Terrier for the experiments
  - System version not reported
- Use standard test collections
  - Bulgarian and French # of docs

Conclusions

- Employment of a standard open-source system as Terrier for the experiments
  - System version not reported
- Use standard test collections
  - Bulgarian and French # of docs
  - WSJ sub-corpus

Conclusions

- Employment of a standard open-source system as Terrier for the experiments
  - System version not reported
- Use standard test collections
  - Bulgarian and French # of docs
  - WSJ sub-corpus
- Report the pseudo-code of the key algorithms is given for SNS and GRAS

Conclusions

- Employment of a standard open-source system as Terrier for the experiments
  - System version not reported
- Use standard test collections
  - Bulgarian and French # of docs
  - WSJ sub-corpus
- Report the pseudo-code of the key algorithms is given for SNS and GRAS
- Some missing information about key parameters and rule-based stemmers employed


ANY
QUESTIONS
?