

CITREC Similarity Tables

The following table describes the naming conventions for database tables containing calculated similarity scores. All tables are prefixed with “sim_”. Prefix and suffix are separated by “_”. E.g. for the Amsler measure, the database tables are named “sim_amsler” (for the regular implementation) and “sim_amsler_rel” (for the normalized version).

Name prefix	Name suffix	Description
amsler		Amsler [1]
	rel	normalized
bibco		Bibliographic Coupling [4]
	rel	normalized
cocit		Co-Citation [6]
	rel	normalized
context		Contextual Cocitation [2]
	avg	averaged score
	pow1, pow2	score weighted with the 1 st / 2 nd power of the Co-Citation strength
	rt2, rt5, rt10	score weighted with the 2 nd / 5 th / 10 th root of the Co-Citation strength
	sum	all scores summed up (eq. to pow1)
cpa		Citation Proximity Analysis [3]
	2simple_tree, 5simple_tree, 10simple_tree	score of Basic CPA weighted 2 / 5 / 10 times plus Extended CPA
	pow1, pow2	Basic CPA score weighted with the 1 st / 2 nd power of the Co-Citation strength
	rt2, rt5, rt10	Basic CPA score weighted with the 2 nd / 5 th / 10 th root of the Co-Citation strength
	simple	Basic CPA score
	simple_2tree, simple_5tree, simple_10tree	score of Basic CPA plus Extended CPA weighted 2 / 5 / 10 times
	simple_tree	score of Basic CPA plus Extended CPA
	sum	all scores summed up (eq. to pow1)
	tree	Basic CPA score
linkthrough		Transitive similarity (required for Amsler)

lucene(b)		Lucene MoreLikeThis ¹ (with Boosts)
	title_abstract	Title (boost 4) and abstract (boost 2)
	title_abstract_text	Title (boost 4), abstract (boost 2) and text
	abstract	Abstract
	text	Text
	title	Title
mesh	lin	Similarity measure by Lin using MeSH [5]

References

- [1] R.A. Amsler. Applications of citation-based automatic classification. Technical report, University at Austin, Linguistics Research Center, Texas, 1972.
- [2] Alison Callahan, Stephen Hockema, and Gunther Eysenbach. Contextual Cocitation: Augmenting Cocitation Analysis and its Applications. *Journal of the American Society for Information Science and Technology*, 61:1130–1143, March 2010.
- [3] Bela Gipp and Joeran Beel. Citation Proximity Analysis (CPA) - A new approach for identifying related work based on Co-Citation Analysis. In Birger Larsen and Jacqueline Leta, editors, *Proceedings of the 12th International Conference on Scientometrics and Informetrics (ISSI'09)*, volume 2, pages 571–575, Rio de Janeiro (Brazil), July 2009. International Society for Scientometrics and Informetrics. ISSN 2175-1935. Available at <http://sciplore.org>.
- [4] M. M. Kessler. Bibliographic coupling between scientific papers. *American Documentation*, 14:10–25, 1963.
- [5] Dekang Lin. An Information-Theoretic Definition of Similarity. In *Proceedings of the 15th International Conference on Machine Learning*, pages 296–304, 1998.
- [6] H Small. Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24:265–269, 1973.

¹http://lucene.apache.org/core/old_versioned_docs/versions/3_5_0/api/all/org/apache/lucene/search/similar/MoreLikeThis.html