

ConceptNet 5

Robert Speer and Catherine Havasi
MIT Media Lab

ABSTRACT

ConceptNet is a knowledge representation project, providing a large semantic graph that describes general human knowledge and how it is expressed in natural language. The scope of ConceptNet includes words and common phrases in any written human language. It provides a large set of background knowledge that a computer application working with natural language text should know. These words and phrases are related through an open domain of predicates, describing not just how words are related by their lexical definitions, but also how they are related through common knowledge.

Transforming ConceptNet into a vector space allows reasoning about similarity and relatedness of words with speed and broad coverage[1]. As ConceptNet currently has data in many written languages, we can build a vector space containing concepts from each of our languages enabling cross-language reasoning similar to parallel corpora in translation. Applications of this technology include free text analytics, sentiment analysis, intelligent search and machine reading.

More information on ConceptNet including its fundamental design decisions, ways to use it, and evaluations of its coverage and accuracy can be found here[2]. Its content is searchable and downloadable under a Creative Commons license from <http://conceptnet5.media.mit.edu>.

BODY

ConceptNet 5 is released. It's a free semantic network in many languages, enabling reasoning and machine learning about what words mean.

REFERENCES

- [1] C. Havasi, R. Speer, J. Pustejovsky, and H. Lieberman. Digital intuition: Applying common sense using dimensionality reduction. *IEEE Intelligent Systems*, 24(4):24–35, July 2009.
- [2] R. Speer and C. Havasi. Representing general relational knowledge in conceptnet 5. In N. C. C. Chair), K. Choukri, T. Declerck, M. U. Dogan, B. Maegaard, J. Mariani, J. Odiijk, and S. Piperidis, editors, *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may 2012. European Language Resources Association (ELRA).

Volume X of Tiny Transactions on Computer Science

This content is released under the Creative Commons Attribution-NonCommercial ShareAlike License. Permission to make digital or hard copies of all or part of this work is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.
CC BY-NC-SA 3.0: <http://creativecommons.org/licenses/by-nc-sa/3.0/>.