

Enhanced Features in the SPECIALIST Lexicon - Antonyms

Chris J. Lu, PhD^{1,2}, Amanda Payne, PhD^{1,2} and James G. Mork, MSc¹

¹National Library of Medicine, Bethesda, MD ²Medical Science & Computing, LLC, Rockville, MD

Introduction

Antonyms are words that have opposite or contrasting meanings in a specific domain [1]. For example, “asleep” is the opposite of “awake” in the domain of “physical property” and they are considered complementary antonyms without middle ground, while “long” is the opposite of “short” in the domain of “size” and they are considered scalar antonyms with levels of comparison. Antonyms are essential for many NLP applications, such as negation detection, paraphrasing, contradiction detecting, question answering, machine translation, sentiment analysis, information retrieval and textual inference. There are no publicly available comprehensive antonym lists. Accordingly, our objective is to develop a systematic approach to generate antonyms in the SPECIALIST Lexicon (thereafter, the Lexicon) and hope to provide generic and comprehensive antonym features needed for the NLP community.

Implementation and Usage

Terms in the Lexicon antonym pairs (aPairs) must be single words in the Lexicon with the same Part-Of-Speech (POS). Terms that meet these criteria are retrieved from two models as antonym candidates. First, the affixal negation model utilizes prefix and suffix derivations with negation tags from the Lexicon [2]. Second, the collocater model retrieves high frequency co-occurrence terms from a corpus (MEDLINE n-gram set) [3] that are not lexicon synonyms [4] nor affixal negations. These candidates are then sent to linguists for further tagging on canonical, type (B: bounded, UB: unbounded, AB: asymmetric bounded) and negation (N: strict negative, BN: broad negative, O: otherwise, not negative). The generated antonyms are generic and can be used in various NLP applications. For example, negation detection cue words can be retrieved from negative antonyms with strict negation (N), such as “unsuccessful”, “useless”, and “without” in Table 1. Negated bounded (B) antonyms can be used to substitute synonymous antonyms for better recall (not awake = asleep or not asleep = awake). The Lexicon is distributed with UMLS by NLM via an Open Source License agreement and is available at: <https://umlslex.nlm.nih.gov/lexicon>.

Table 1. Examples of antonym list with POS, type, negation, domain and candidate model

Antonym-1 (positive)	Antonym-2 (negative)	POS	Type	Negation	Domain	Candidate Model
successful	unsuccessful	adj	UB	N	quality	affixal negation - prefix
careful	careless	adj	UB	BN	quality	affixal negation - suffix
useful	useless	adj	UB	N	quality	affixal negation - suffix
asleep	awake	adj	B	O	physical property	collocates in a corpus
long	short	adj	UB	O	size	collocates in a corpus
good	bad	adj	AB	BN	quality	collocates in a corpus
with	without	prep	B	N	existence	collocates in a corpus

Acknowledgements

This work was supported by the intramural research program at the U.S. National Library of Medicine, National Institutes of Health.

References

1. Tesfaye D and Paradis C. On the use of antonyms and synonyms from a domain perspective. The NetWordsS Final Conference, Pisa, March 30-April 1, 2015; 150-4.
2. Lu CJ, McCreedy L, Tormey D, et al. A Systematic Approach for Automatically Generating Derivational Variants in Lexical Tools Based on the SPECIALIST Lexicon. IEEE IT Professional Magazine, May/June, 2012; 36-42.
3. Lu CJ, Tormey D, McCreedy L, et al. Generating A Distilled N-Gram Set: Effective Lexical Multiword Building in the SPECIALIST Lexicon. BIOSTEC, HEALTHINF 2017; 5:77-87.
4. Lu CJ, Tormey D, McCreedy L, et al. Enhanced LexSynonym Acquisition for Effective UMLS Concept Mapping. MedInfo 2017, Hangzhou, China, August 21-25, 2017; 245:501-5.