Abstract

Although, the fair amount of works in sentiment analysis (SA) and opinion mining (OM) systems in the last decade and with respect to the performance of these systems, but it still not desired performance, especially for morphologically-Rich Language (MRL) such as Arabic, due to the complexities and challenges exist in the nature of the languages itself. One of these challenges is the detection of idioms or proverbs phrases within the writer text or comment. An idiom or proverb is a form of speech or an expression that is peculiar to itself. Grammatically, it cannot be understood from the individual meanings of its elements and can yield different sentiment when treats as separate words. Consequently, In order to facilitate the task of detection and classification of lexical phrases for automated SA systems, this paper presents AIPSeLEX a novel idioms/proverbs sentiment lexicon for modern standard Arabic (MSA) and colloquial. AIPSeLEX is manually collected and annotated at sentence level with semantic orientation (positive or negative). The efforts of manually building and annotating the lexicon are reported. Moreover, we build a classifier that extracts idioms and proverbs, phrases from text using n-gram and similarity measure methods. Finally, several experiments were carried out on various data, including Arabic tweets and Arabic microblogs (hotel reservation, product reviews,
and TV program comments) from publicly available Arabic online reviews websites (social media, blogs, forums, e-commerce web sites) to evaluate the coverage and accuracy of AIPSeLEX.

References

Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA),
Atlanta, Georgia, 2013, pp. 55–64.
- E. Refaee and V. Rieser, "An Arabic Twitter Corpus for Subjectivity and Sentiment
- A. S. Zibin and A. R. M. Altakhaineh, "Informativity of Arabic Proverbs in
Context: An Insight into Palestinian Discourse," International Journal of Linguistics (IJL),
vol. 6, p. 67, 2014.
- Z. Kovacs, Metaphor:A Practical Introduction, First ed.: Oxford University Press,
2002.
- H. S. Ibrahim, S. M. Abdou, and M. Gheith, "Automatic expandable large-scale
sentiment lexicon of Modern Standard Arabic and Colloquial," in 16th International
Conference on Intelligent Text Processing and Computational Linguistics (CICLING), Cairo -
Egypt, 2015.
- M. Elmahdy, G. Rainer, M. Wolfgang, and A. Slim, "Survey on common Arabic
language forms from a speech recognition point of view," in proceeding of International
- A. Basha, ?????? ???????: ?????? ?????? ???? ?????? ?? ????? ?? ????
?????? [Colloquial sayings: an annotated and arranged by the first letter of ideals with the
Scout TOPICAL]. Egypt: Al-Ahram Foundation - Al-Ahram Center for Translation and
- A. Saalan, ?????? ??????? ??????? ??????? [Encyclopedia of Egyptian popular
- F. Husain, ?????? ???????, ?????? ???????, ??????? ???????, ?????????? ???????
[Colloquial sayings, Egyptian folklore]. Egypt: General Egyptian Book Organization GEBO,
1984.
- PROz. (2014). PROz website for Arabic Idioms/Maxims/Sayings (Jan 2014). Available:
http://www. proz. com/glossary-translations/
- J. C. Carletta, "Assessing agreement on classification tasks: the KAPPA statistic
- G. Salton, A. Wong, and C. S. Yang, "A vector space model for information
- V. Levenshtein, "Binary codes capable of correcting deletions, insertions and
- X. Ding, B. Liu, and P. S. Yu, "A holistic lexicon-based approach to opinion
mining," in Proceeding WSDM &apos;08 Proceedings of the 2008 International
Conference on Web Search and Data Mining, New York, NY, USA, 2008, pp. 231-240.
- A. Mudinas, D. Zhang, and M. Levene, "Combining lexicon and learning based
approaches for concept-level sentiment analysis," in Proceeding WISDOM &apos;12

Index Terms

Computer Science

Information Sciences

Keywords

Sentiment lexicon  Idioms lexicon  AIPSeLEX.