

Title	Overview of tools for sentiment analysis and opinion mining
Version	1.0
Author(s)	Darja Fišer, Jakob Lenardič
Date	22-07-2020
Status	Draft
Distribution	BoD, NCF, UI
ID	CE-2020-1640

Table of contents

1. Introduction.....	1
2. Tools for sentiment analysis/opinion finding in the CLARIN infrastructure.....	2
2.1. The tools.....	2
3. Overview of the tools.....	4
3.1. Identification.....	4
3.2. Availability.....	4
3.3. Metadata.....	4
3.3.1. Language.....	4
3.3.2. Functionality and sentiment type.....	4
3.3.3. Domain.....	5
3.3.4. Licence.....	5
4. Conclusion.....	5
5. References.....	6

1. Introduction

In this report, we present an overview of language tools dedicated to sentiment analysis and opinion mining. These are text analysis methods that identify and extract people’s opinions, attitudes and sentiments within a text (Zhang and Liu 2017, 1153). Although the two terms are often (and confusingly) used interchangeably (e.g., [Wikipedia](#)), we distinguish between the two and use the term *sentiment analysis* to refer to the method of determining the sentiment of a particular sentence (or potentially other grammatical construction) by assigning a ternary (“positive”, “neutral”, “negative”) or scalar value to the sentence. By contrast, *opinion mining* is a text extraction method targeting those parts of a text in which a person’s attitude, opinion, or sentiment is expressed. We do this because the sentiment analysis/opinion mining tools listed in the CLARIN infrastructure generally distinguish between the two functionalities as well, as will be shown in Section 2.1. In terms of applicability, sentiment analysis and opinion mining are widely used in domains like social media or customer reviews, where the user’s voice is expressed.

The overview was conducted in two steps:

- (i) We added sentiment analysis tools listed in the [Virtual Language Observatory \(VLO\)](#). The VLO was searched with the keywords *sentiment (analysis) tool/software, opinion (finding/mining), sentiment analyser, and sentiment analyzer*.
- (ii) Afterwards, we asked national CLARIN UI coordinators to provide additional information on sentiment analyzers from their own countries in a [Google Docs survey](#).¹

In this way we collected a small set of 5 tools dedicated to sentiment analysis or opinion minding. Our primary aim of this survey was to evaluate the presentation of their availability and metadata (primarily language, scope of sentiment analysis, and licence).

2. Tools for sentiment analysis/opinion finding in the CLARIN infrastructure

2.1. The tools

Table 1: The tools for sentiment analysis/opinion finding, sorted by language

Tool	Languages	Description
Etuma Customer Feedback Analysis Functionality: sentiment analysis, opinion mining Domain: customer feedback Licence: under negotiation (commercial)	English, French, Italian, German, Swedish, Norwegian, Danish, Finnish, Spanish, Dutch	The tool is used for ranking customers' feedback in order of intensity and detects the sentiment of ongoing discussions in order to determine whether the overall response to a product or campaign is positive or negative. CLARIN centre: FIN-CLARIN
finsentiment Functionality: sentiment analysis Domain: product review, social media	Finnish	This tool relies on three resources: <ol style="list-style-type: none"> (i) word embeddings calculated from a corpus of Finnish text; (ii) product reviews harvested from the Internet; (iii) a word-based convolutional neural network with 100 kernels each of sizes 2, 3, 4 and 5 words. The neural network is trained to predict the rating associated with product reviews, and the prediction it gives to the input text is converted to a sentiment.

¹ We would like to thank all the UI representatives and National Coordinators who have participated in the survey.

		<p>Availability: online</p> <p>Sentiment type: document and sentence-level sentiment</p> <p>CLARIN centre: FIN-CLARIN</p>
<p>OptaHopper: phrase-level sentiment with opinion targets</p> <p>Functionality: sentiment analysis, opinion mining</p> <p>Domain: independent</p> <p>Licence: BSD 2 Clause</p>	Polish	<p>This is a phrase- and sentence-level sentiment analysis tool based on TreeLSTM (Sheng Tai et al. 2015) integrated with opinion mining. Any sentiment dictionary may be used as an input feature, including lemma-level and plWordNet emo. In the case of plWordNet emo, provided integration with the WSD module. The Opinion Finder OPFI (Wawer 2016) can be used for opinion target extraction.</p> <p>Availability: download</p> <p>Sentiment type: phrase- and sentence-level sentiment; opinion mining</p> <p>CLARIN Centre: CLARIN-PL</p>
<p>Sentiment Analysis Tool</p> <p>Functionality: sentiment analysis, opinion mining</p> <p>Domain: independent</p> <p>Licence: LGPL</p>	Greek	<p>The sentiment analysis tool is a text classification and sentiment extraction tool based on n-gram graph text representations. It may be paired with various machine learning algorithms for the generation of the language model. It can be accessed by a URL endpoint as a REST service. It has been used as is, or as part of bigger pipelines in many research tasks. It is also embedded in gov.insight as an annotator producer for sentiment classification.</p> <p>This tool also makes use of an Opinion Mining process.</p> <p>Availability: online</p> <p>CLARIN centre: CLARIN:EL</p> <p>Publication: Kiomourtzis et al. (2014)</p>
<p>TreeHopper (TreeLSTM): wydźwięk na poziomie zdań i fraz</p> <p>Functionality: sentiment analysis</p> <p>Domain: independent</p>	Polish	<p>This is a TreeLSTM-based (Sheng Tai et al. 2015) dependency tree sentiment labeller, implemented in PyTorch and optimized for morphologically rich languages with a relatively loose word order (such as Polish).</p> <p>Availability: download</p>

Licence: BSD 2 Clause		Sentiment type: dependency subtree/phrase sentiment CLARIN Centre: CLARIN-PL Publication: Korbak and Żak (2017)
------------------------------	--	---

3. Overview of the tools

3.1. Identification

2 out of the 5 tools cannot be found in the VLO:

- (i) [finsentiment](#) (FIN-CLARIN)
- (ii) [Sentiment Analysis Tool](#) (CLARIN:EL)

3.2. Availability

The following 2 tools are available for download, both from the CLARIN-PL repository:

- (i) [OptaHopper: phrase-level sentiment with opinion targets](#)
- (ii) [TreeHopper \(TreeLSTM\): wydźwięk na poziomie zdań i fraz](#)

The following 2 tools are available for online. Tool (i) is available as a demo through the Finnish Language Bank, while tool (ii) is directly integrated as a webservice through CLARIN:EL.

- (i) [finsentiment](#)
- (ii) [Sentiment Analysis Tool](#)

Lastly, [Etuma Customer Feedback Analysis](#) is a commercial tool and thus isn't freely available.

3.3. Metadata

3.3.1. Language

4 out of the 5 tools can be used for sentiment analysis/opinion mining within a single language:

- (i) Polish (2 tools) – [OptaHopper: phrase-level sentiment with opinion targets](#) and [TreeHopper \(TreeLSTM\): wydźwięk na poziomie zdań i fraz](#)
- (ii) Greek (1 tool) – [Sentiment Analysis Tool](#)
- (iii) Finnish (1 tool) – [finsentiment](#)

The remaining tool – [Etuma Customer Feedback Analysis](#) – has a multilingual scope and can be used for opinion mining in 10 languages: English, French, Italian, German, Swedish, Norwegian, Danish, Finnish, Spanish, Dutch.

3.3.2. Functionality and sentiment type

The tools in Table 1 have the following functionality:

- (i) [OptaHopper: phrase-level sentiment with opinion targets](#)
 - a. *Phrase-level sentiment*
 - b. *Sentence-level sentiment*
 - c. *Opinion mining*

- (ii) [TreeHopper \(TreeLSTM\): wydźwięk na poziomie zdań i fraz](#)
 - a. *Dependency subtree sentiment*
 - b. *Phrase-level sentiment*
- (iii) [Etuma Customer Feedback Analysis](#)
 - a. *Sentence-level sentiment*
 - b. *Opinion mining*
- (iv) [finsentiment](#)
 - a. *Sentence-level sentiment*
 - b. *Document-level sentiment*
- (v) [Sentiment Analysis Tool](#)
 - a. *Sentence-level sentiment*
 - b. *Opinion mining*

In sum, all the tools are used for the analysis of sentiment at least at the sentence level. 3 out of the 5 tools can also be used for opinion mining.

It is noteworthy that both CLARIN-PL tools – i.e., [OptaHopper: phrase-level sentiment with opinion targets](#) and [TreeHopper \(TreeLSTM\): wydźwięk na poziomie zdań i fraz](#) – determine the sentiment below the sentence level, i.e., that of phrasal constituents, so they take syntactically parsed data as input.

3.3.3. Domain

The following 3 tools are domain independent:

- (i) [OptaHopper: phrase-level sentiment with opinion targets](#)
- (ii) [TreeHopper \(TreeLSTM\): wydźwięk na poziomie zdań i fraz](#)
- (iii) [Sentiment Analysis Tool](#)

The following 2 tools are used within a particular domain (listed in the parentheses):

- (i) [Etuma Customer Feedback Analysis](#) (customer feedback)
- (ii) [finsentiment](#) (product review, social media)

3.3.4. Licence

The only tool that lacks information about its licence is [finsentiment](#), while the licence of [Etuma Customer Feedback Analysis](#) is listed as under negotiation in META-SHARE (FIN-CLARIN distribution). Otherwise, the following licences are represented:

- (i) BSD 2 Clause (2 tools)
- (ii) LGPL (1 tool)

4. Conclusion

Only 5 tools for sentiment analysis/opinion mining are available through the CLARIN infrastructure, which is few in comparison to the tools surveyed in the previous reports (e.g., there are 24 tools for [Named Entity Recognition](#)). 3 out of the 5 tools can be found in the VLO.

In terms of availability, 2 tools are available for download, both from the CLARIN-PL repository, while 2 tools are available as a webservice. In terms of the language, 4 out of 5 tools have a monolingual scope (2 tools for Polish, 1 tool for Greek, and 1 tool for Finnish).

In relation to functionality, all tools detect sentiment at the sentiment levels. 3 out of the 5 tools can also be used for opinion mining, while the 2 CLARIN-PL tools can also determine the sentiment of phrasal constituents. 3 tools are domain independent, while 1 tool is used for sentiment analysis/opinion mining in the domain of customer feedback and 1 tool is used to determine sentiment/mine opinions in the domains of social media and product review.

All tools except for one contain information on licence.

5. References

- Kiomourtzis, George, George Giannakopoulos, Georgios Petasis, Pythagoras Karampiperis, and Vangelis Karkaletsis. 2014. NOMAD: Linguistic Resources and Tools Aimed at Policy Formulation and Validation. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, edited by Nicoletta Calzolari et al., 26–31. European Language Resources Association. <https://www.aclweb.org/anthology/L14-1629/>.
- Korbak, Tomasz, and Paulina Żak. 2017. Fine-tuning Tree-LSTM for phrase-level sentiment classification on a Polish dependency treebank. Submission to PolEval task 2. <https://arxiv.org/abs/1711.01985>.
- Sheng Tai, Kai, Richard Socher, and Christopher D. Manning. 2015. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, edited by Chengqing Zong and Michael Strube, 1556–1566. Association for Computational Linguistics. <http://dx.doi.org/10.3115/v1/P15-1150>.
- Wawer, Aleksander. 2016. OPFI: A Tool for Opinion Finding in Polish. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, edited by Nicoletta Calzolari et al., 2906–2909. European Language Resources Association. <https://www.aclweb.org/anthology/L16-1464>.
- Zhang, Lei, and Bing Liu. 2017. Sentiment Analysis and Opinion Mining. In *Encyclopedia of Machine Learning and Data Mining*, edited by Claude Sammut and Geoffrey I. Webb, 1152–1161. https://doi.org/10.1007/978-1-4899-7687-1_907.