

TURKSENT: A Sentiment Annotation Tool for Social Media

Gülşen Eryiğit

Dep. of Computer Eng.
Istanbul Technical University
gulsen.cebiroglu@itu.edu.tr

Fatih Samet Çetin and Meltem Yanık

Dep. of Information Technology
Turkcell Global Bilgi
fatih.cetin@global-bilgi.com.tr
meltem.yanik@global-bilgi.com.tr

Tanel Temel

Dep. of Information Technology
Turkcell Global Bilgi
tanel.temel@global-bilgi.com.tr

İlyas Çiçekli

Dep. of Computer Eng.
Hacettepe University
ilyas@cs.hacettepe.edu.tr

Abstract

In this paper, we present an annotation tool developed specifically for manual sentiment analysis of social media posts. The tool provides facilities for general and target based opinion marking on different type of posts (i.e. comparative, ironic, conditional) with a web based UI which supports synchronous annotation. It is also designed as a SaaS (Software as a Service). The tool's outstanding features are easy and fast annotation interface, detailed sentiment levels, multi-client support, easy to manage administrative modules and linguistic annotation capabilities.

1 Introduction

Today, monitoring social media is a vital need for companies and it has a high commercial value. So almost all companies have social media accounts and departments for following the social media about their business sectors. In recent decade, the studies on sentiment analysis has gained high popularity and several academic (Pang and Lee, 2007; Liu, 2012) and commercial (Radian6, 2013; Lithium, 2013) projects emerged in this field. Although there are many works (Bosco et al., 2013; Wiebe et al., 2005) on creating sentiment corpora, up to our knowledge there are no publicly available and professional sentiment annotation tools.

A huge contact center communicates with the customers for different trade marks on behalf of them and provides detailed CRM¹, impact

and competitor analysis reports. With this purpose, they employ thousands of customer representatives among which an increasing percentage should deal with social media monitoring, the new channel of communication. In such an environment, the monitoring should be done via professional and synchronous UIs (user interfaces) where the performance of each human agent has high importance. Most of the current commercial monitoring tools lacks the following features:

- a detailed sentiment analysis interface for feature based and comparative opinion declarations,
- an effective and synchronous annotation interface,
- on-demand data loading,
- linguistic annotation modules,
- detailed data analyses for corpus creation (to be used in supervised machine learning).

The aim of our work is to fulfill all of the above listed requirements and provide a platform for effective annotation of social media data. The tool has the following sentiment and linguistic annotation layers:

- general and target based sentiment
- text normalization
- named entity
- morphology
- syntax

The sentiment annotation module of TURKSENT may operate multilingually whereas the linguistic annotation module is initially configured

¹CRM: Customer Relationship Management

specific to Turkish following the work in ITU Treebank Annotation Tool (Eryigit, 2007). It is also possible to adapt this part to other languages by plugging relevant linguistic adapters (for semi-automatic annotation).

TURKSENT will be freely available for academic projects as a SaaS.

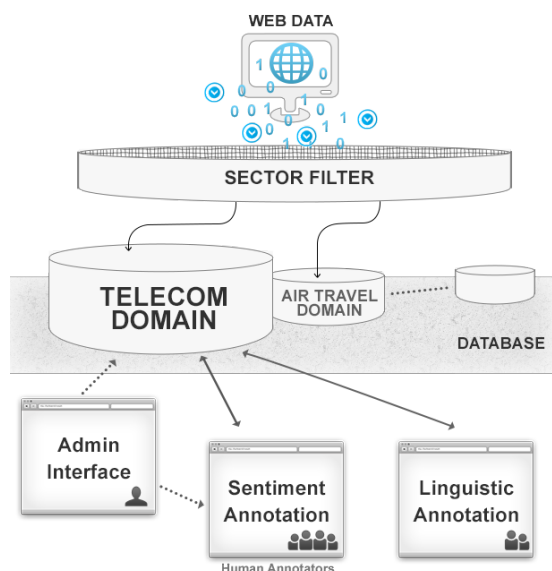


Figure 1: Application Flow

2 Architecture

Figure 1 gives an idea about the flow of our application. In our system, the web data is monitored continuously. It is first of all filtered according to the target sector by the “sector filter” and it is then stored in the relevant database domains. In our system, each domain represents a workspace which consists of the related sector data (collected via web or uploaded manually to the system), an administrator and a bunch of human annotators.

2.1 Sentiment Annotation

Our choice of SaaS design has the following goals:

- Platform independence (No special machine or no special operating system)
- Accessibility (Accessible from anywhere anytime by multiple users)
- No installation effort (Browser based application)

- No need to deploy updates to clients

Figure 2 gives a sample sentiment annotation screen-shot on an example Tweet (“Samsung Galaxy S4’s hardware features are amazing but software is not stable as Iphone”). The upper half of the screen (up to the table) show the general sentiment part which is tagged as *both*² (the ambivalent smiley). General sentiment tagging means identifying the sentimental class regardless of a target. In other words, extracting dominant sentimental class of an instance. In this stage the annotator is also expected to select an appropriate comment category and sentence type.

The lower half is for target based sentiment annotation. These deep sentiments are represented as tuples consisting of the brand, product/service, feature and sentiment tags. For example, the first tuple in the sample Tweet will be composed as the following: <Samsung, Galaxy S4, hardware, positive> which means the hardware feature of the Samsung Brand’s product Galaxy S4 had a positive impact on the Tweet’s author.

2.2 Linguistic Annotation

Recent researches on sentiment analysis show that it is not possible to really understand the sentiment of a sentence without any natural language processing (NLP). And the addition of NLP features to these systems increases the success ratios of the automatic analyzers dramatically. In order to be able to prepare a sentiment corpus, being able to annotate the focus data within the same platform is an important issue. Furthermore, the web data has severe differences when compared to formal natural language text and it needs additional preprocessing before linguistic phases. With this need, we added a linguistic annotation interface to our application which is basically a reimplementation and adaptation of a previous academic study (Eryigit, 2007) according to our needs.

In this layer, the linguistic expert annotator is asked to first normalize the instances (i.e. misspellings, exaggerations, web jargon), and then determine the entities (ex: “Galaxy S4”), select the appropriate postag categories for words and annotate the syntactic parts of a post. It is also possible to operate this layer semi-automatically by using the pretrained linguistic tools and outputting their

²Other options are: positive, negative and neutral(no sentimental expression at all).

TURKSENT ANNOTATION TOOL
Welcome, Mehmet Osmanoglu

Sentiment Analysis
Morphological Disambiguation
Dependency Parsing

Sentiment Analysis

Telekom

Skip F9
Skipped Posts

Post ID: 754

Samsung Galaxy S4's hardware features are amazing but software is not stable as iPhone's.

Comment Categories

Satisfaction Complaint

Sentence Types

Comparative

	Brand	Product / Service	Feature	Sentiment	Reason
1	Samsung	Galaxy S4	Hardware		amazing
2	Samsung	Galaxy S4	Software		not stable
3	Apple	iPhone	Software		
4					

Figure 2: Sentiment annotation

results to the human experts and taking their corrections. This speed-up procedure is only available for Turkish now, but the tool is developed as a pluggable architecture to support further studies on other languages. Figure 3 shows some sample screenshots for the linguistic layer.

2.3 Administrative Operations

TURKSENT has a simple and easy-to-use admin interface. A user who has administration rights has the ability to perform the actions listed below:

- Creating a workspace (with a focus data and annotator group)
- Determining the data subsets for linguistic annotation
- Controlling/Changing the ongoing annotations
- Defining configurable items (sentence types, comment categories, product/service list, feature list, brand list)
- Defining linguistic tags (pos tags, named entity types, dependency types)

3 Usability

The usability is seriously taken into account during the design and development of our application. The spent time per post is a high concern within big operations. End-user software tests are accomplished and observed for each step. On the final UI design, every action can be done via keyboard without the need of mouse usage. Almost every text areas has strong auto-completion feature in itself. While an annotator is working on an issue, it is possible to deliver any idea-suggestion to the administrator within seconds. And if an annotator need to browse his/her previous annotations, can easily search and find within them.

4 Conclusion

In this work, we presented a professional sentiment annotation tool TURKSENT which supports synchronous annotations on a web-based platform. The study is a part of an automatic sentiment analysis research project. That is why, it both aims to manually annotate the sentiments of web posts and to create a sentiment corpus also annotated linguistically (to be used in automatic

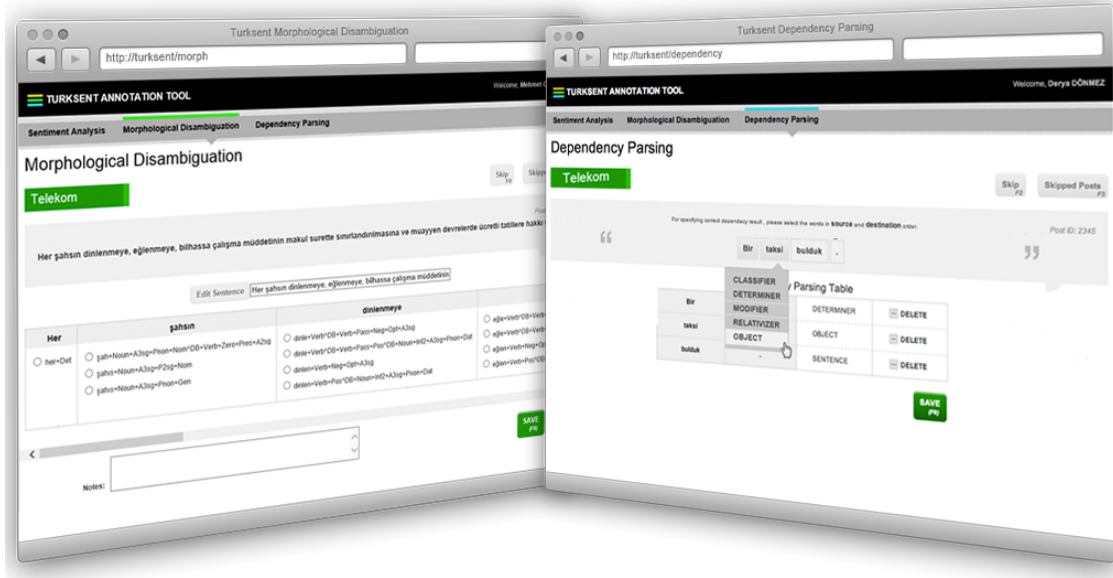


Figure 3: Linguistic Annotations

sentiment analysis). With this purpose it consists different layers of annotation specific to web data. It serves as a SaaS and designed as dynamic as possible for future use on different sectors and languages.

Acknowledgment

This work is accomplished as a part of a TUBITAK-TEYDEB (The Scientific and Technological Research Council of Turkey - Technology and Innovation Funding Programs Directorate) project (grant number: 3120605) in “Turkcell Global Bilgi” Information Technology Department. The authors want to thank Derya Dönmez and Mehmet Osmanoğlu for design and implementation.

References

- Cristina Bosco, Viviana Patti, and Andrea Bolioli. 2013. Developing corpora for sentiment analysis and opinion mining: the case of irony and senti-tut. *Intelligent Systems*.
- Gülşen Eryiğit. 2007. ITU Treebank Annotation Tool. In *Proceedings of the ACL workshop on Linguistic Annotation (LAW 2007)*, Prague, 24-30 June.
- Lithium. 2013. Lithium. <http://www.lithium.com/>.

Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan and Claypool Publishers.

Bo Pang and Lillian Lee. 2007. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1-135.

Radian6. 2013. Radian 6. <http://www.salesforcemarketingcloud.com/products/social-media-listening/>.

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 39(2-3):165-210.