

ESWC-16 Challenge on Fine-Grained Sentiment Analysis

The Fine-Grained Sentiment Analysis Challenge is open to everyone from industry and academia working within the sentiment analysis area.

Background and Relevance for the Semantic Web community

As the Web rapidly evolves, Web users are evolving with it. In an era of social connectedness, people are becoming increasingly enthusiastic about interacting, sharing, and collaborating through social networks, online communities, blogs, Wikis, and other online collaborative media. In recent years, this collective intelligence has spread to many different areas, with particular focus on fields related to everyday life such as commerce, tourism, education, and health, causing the size of the Social Web to expand exponentially.

The opportunity to automatically capture the opinions of the general public about social events, political movements, company strategies, marketing campaigns, and product preferences has raised growing interest within the scientific community, leading to many exciting open challenges, as well as in the business world, due to the remarkable benefits deriving from marketing prediction. The distillation of knowledge from such a large amount of unstructured information is an extremely difficult task, as the contents of today's Web are perfectly suitable for human consumption, but remain hardly accessible to machines. Mining opinions and sentiments from natural language, involves a deep understanding of most of the explicit and implicit, regular and irregular, syntactical and semantic rules proper of a language. Existing approaches mainly rely on parts of text in which opinions and sentiments are explicitly expressed such as polarity terms, affect words and their co-occurrence frequencies. However, opinions and sentiments are often conveyed implicitly through latent semantics, which make purely syntactical approaches ineffective. This issue offers a research opportunity and

an exciting challenge to the Semantic Web community. In fact, fine-grained sentiment analysis aims to go beyond a mere word-level analysis of text and provide novel approaches to opinion mining and sentiment analysis that allow a more efficient passage from (unstructured) textual information to (structured) machine-processable data, in potentially any domain. Besides, there might be implicit opinions deriving from particular use of verbs that are difficult to catch from a sentiment analysis tool.

As an example, a sentence such as *People hope that the President will be condemned by the judges* includes a triggering event *hope* that carries a positive sentiment tone of the expressed opinion. However, this is not enough for a complete sentiment analysis of this sentence. With classical sentiment analysis approaches we can only state that the sentence expresses a positive opinion on the event *condemn* and nothing else.

With fine-grained sentiment analysis we can go one step deeper as it focuses on a semantic analysis of text through the use of web ontologies, semantic resources, or semantic networks, allowing the identification of opinion data which with only natural language techniques would be very difficult. Fine-Grained sentiment analysis allows, in the example above, detecting a negative opinions of the holder *People* towards the subject *President*. By relying on large semantic knowledge bases, fine-grained sentiment analysis steps away from blind use of keywords and word co-occurrence count, but rather relies on the implicit features associated with natural language concepts. Unlike purely syntactical techniques, concept-based approaches are able to detect also sentiments that are expressed in a subtle manner, e.g., through the analysis of concepts that do not explicitly convey any emotion, but which are implicitly linked to other concepts that do so.

Potential attendance

This challenge would be the third edition of two successful events held at ESWC 2014 and 2015. In 2014 the challenge had six successful submissions (10 intentional but not accepted because the proposed semantic engine did not look effective) and in 2015 it had four successful submissions (four intentional). In both the editions all teams attended the conference with at least one representative. Most of them came only for the challenge, i.e. they did not have any other commitment or contribution at the conference. In the

forthcoming 2016 edition, we expect to increase this number significantly as we have the opportunity to exploit lessons learnt from last years experience, participants' feedback and a promotion of the challenge (performed much before the submission of this proposal) to all the previous challengers of the Sentiment Analysis track of the SemEval workshop and of our previous editions of the challenge. In fact we have already extracted information (name, affiliation and email) of 163 previous mentioned participants and have already sent them the list of the tasks mentioned in this proposal asking their interest in participating in the proposed challenge. We received the following feedbacks: **X** of them will participate, **Y** are interested but uncertain to participate and **Z** did not reply. Moreover, similar input and output formats used at the SemEval 2015 Workshop will be adopted here. This solution has a twofold positive effect:

1. most of the teams potentially interested in this challenge usually participate in SemEval;
2. viceversa, having the same input format as SemEval may increase the chance of bringing teams from that community - that usually do not attend ESWC - to submit to our challenge and consequently join ESWC community.

A proposal for a related workshop where challengers can submit their research work (a bit more theoretical and showing more evaluation results) will be submitted at the upcoming edition of the ESWC conference as well. In case of acceptance of the workshop, the challenge will have an additional strength provided by the mutual support between the two events: discussing open issues and research directions at the workshop and showing the last technological achievements at the challenge session. This would further stimulate and encourage participants to present their work at one of the two events.

Call for contributions: essentials

A call for contributions will be prepared according to ESWC challenge track guidelines. It will require participant to submit their system as well as a paper describing their approach, which will undergo a peer-review process. Submissions must include:

1. Abstract: no more than 200 words.

2. Paper: containing the details of the system, including why the system is innovative, which features or functions the system provides, what design choices were made and what lessons were learned, how the semantics has been employed and which tasks the system addresses. The paper should be 6-8 pages length.
3. Web Access: applications should be either accessible via the web or downloadable. If an application is not publicly accessible, password must be provided for reviewers. A short set of instructions on how to use the application should be provided as well.
4. If accepted, the authors will have the possibility to present a poster and a demo advertising their work or networking during a dedicated session.

Possible Important Dates

- **March 3, 2016, 23:59 CET:** Submission due
- **April 9, 2016, 23:59 CET:** Notification
- **May 31-4, 2016:** The Challenge takes place at ESWC-16

Challenge Criteria

This challenge focuses on the introduction, presentation, and discussion of novel approaches to semantic sentiment analysis. Participants will have to design a semantic opinion-mining engine that exploits Semantic Web knowledge bases, e.g., ontologies, DBpedia, etc., to perform multi-domain sentiment analysis. The main motivation for this challenge is to go beyond a mere word-level analysis of natural language text and provide novel semantic tools and techniques that allow a more efficient passage from (unstructured) natural language to (structured) machine-processable data in potentially any domain.

The submitted systems must provide an output according to Semantic Web standards (RDF, OWL, etc.). Systems must have a semantic flavor (e.g., by making use of Linked Data or known semantic networks within their core functionalities) and authors need to show how the introduction of

semantics improves the precision of their methods. Existing natural language processing methods or statistical approaches can be used too as long as the semantics plays a role within the core approach and improves the precision (engines based merely on syntax/word-count will be excluded from the competition). The target language is English and multi-language capability is a plus.

Tasks

The Concept-Level Sentiment Analysis Challenge is defined in terms of different tasks. The first task is elementary whereas the others are more advanced.

Task #1: Polarity Detection

The basic task of the challenge is binary polarity detection. The proposed semantic opinion-mining engines will be assessed according to precision, recall and F-measure of detected polarity values (positive OR negative) for each review of the evaluation dataset. The problem of subjectivity detection is not addressed within this Challenge, hence participants can assume that there will be no neutral reviews. The output format for such a task is the following:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<Sentences>
  <sentence id="apparel_0">
    <text>
      GOOD LOOKING KICKS IF YOUR KICKIN IT OLD SCHOOL LIKE ME. AND COMFORTABLE. AND
      RELATIVELY CHEAP. I'LL ALWAYS KEEP A PAIR OF STAN SMITH'S AROUND FOR WEEKENDS
    </text>
    <polarity>
      1
    </polarity>
  </sentence>
  <sentence id="apparel_1">
    <text>
      These sunglasses are all right. They were a little crooked, but still cool..
    </text>
    <polarity>
      1
    </polarity>
  </sentence>
</Sentences>
```

Input is the same without the polarity tag.

Task #2: Aspect-Based Sentiment Analysis

The output of this Task will be a set of aspects of the reviewed product and a binary polarity value associated to each of such aspects. So, for example, while for the Task #1 an overall polarity (positive or negative) is expected for a review about a mobile phone, this Task requires a set of aspects (such as ‘speaker’, ‘touchscreen’, ‘camera’, etc.) and a polarity value (positive OR negative) associated with each of such aspects. Engines will be assessed according to both aspect extraction and aspect polarity detection using precision, recall and F-measure similarly as performed during the first Concept-Level Sentiment Analysis Challenge held during ESWC2014 and re-proposed at SemEval 2015 Task12¹. The output format for such a task is the following:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
  <Review rid="1">
    <sentences>
      <sentence id="348:0">
        <text>Most everything is fine with this machine: speed, capacity, build.</text>
        <Opinions>
          <Opinion aspect="MACHINE" polarity="positive"/>
        </Opinions>
      </sentence>
      <sentence id="348:1">
        <text>The only thing I don't understand is that the resolution of the
          screen isn't high enough for some pages, such as Yahoo!Mail.
        </text>
        <Opinions>
          <Opinion aspect="SCREEN" polarity="negative"/>
        </Opinions>
      </sentence>
      <sentence id="277:2">
        <text>The screen takes some getting use to, because it is smaller
          than the laptop.</text>
        <Opinions>
          <Opinion aspect="SCREEN" polarity="negative"/>
        </Opinions>
      </sentence>
    </sentences>
  </Review>
```

Input is the same without the Opinions tag and its descendants nodes.

¹www.alt.qcri.org/semEval2015/task12/

Task #3: Semantic Sentiment Retrieval

The output of this Task will be a list of entities ordered by strength of positive judgements of any of their features. As an example, given an input list of reviews on smartphones, create a structured output of each review where smartphone are listed together with their features and opinions on each of them. This task includes Information Retrieval (detect features of given entities), Named Entity Recognition (detect smartphone models within the review possibly using some sort of knowledge base), Sentiment Analysis (aggregate features opinions for the entity sentiment for either overall or feature based retrieval). The input format for such a task is the following:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<Sentences>
  <sentence id="0">
    <text>So far so good. My wife just loves the new Samsung S5: the display is awesome
    and the colors are very brilliant. However, further memory is necessary for storing
    everything.</text>
  </sentence>
  <sentence id="1">
    <text>All the LG G3 have problems with videos: they often are not able to connect
    with tv and when they can, the quality of the image is poor. The only strong point
    is the amount of memory coming from the factory.</text>
  </sentence>
</Sentences>
```

The output format should look like the following:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<Ranks>
  <rank quality="display">
    <position value="1" name="Samsung Galaxy S5"/>
    <position value="2" name="LG G3"/>
  </rank>
  <rank quality="memory">
    <position value="1" name="LG G3"/>
    <position value="2" name="Samsung Galaxy S5"/>
  </rank>
  <rank quality="GENERAL">
    <position value="1" name="Samsung Galaxy S5"/>
    <position value="2" name="LG G3"/>
  </rank>
</Ranks>
```

Task #4: Frame entities Identification

The Challenge focuses on semantic fine-grained sentiment analysis. This means that the proposed engines must work beyond word/syntax level, hence

addressing a concepts/semantics perspective. This task will evaluate the capabilities of the proposed systems to identify the objects involved in a typical opinion frame according to their role: holders, topics, opinion concepts (i.e. terms referring to highly polarised concepts). For example, in a sentence such as *The mayor is loved by the people in the city, but he has been criticized by the state government*², an approach should be able to identify that *the people* and *state government* are the opinion holders, *is loved* and *has been criticized* represent the opinion concepts, *mayor* identifies a topic of the opinion and that there are two different opinion polarities mentioned in the sentence. The proposed engines will be evaluated according to precision, recall and F-measure. The output format for such a task is the following:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<Sentences>
  <sentence id="348:0">
    <text>The mayor is loved by the people in the city,
    but he has been criticized by the state government.
    </text>
    <Frames>
      <Frame>
        <holder start="22" end="32" value="the people"/>
        <topic start="0" end="9" value="The mayor"/>
        <opinion start="10" end="18" value="is loved"/>
        <polarity>positive</polarity>
      </Frame>
      <Frame>
        <holder start="76" end="96" value="the state government"/>
        <topic start="0" end="9" value="The mayor"/>
        <opinion start="53" end="72" value="has been criticized"/>
        <polarity>negative</polarity>
      </Frame>
    </Frames>
  </sentence>
</Sentences>
```

Input is the same without the Frames tag and its descendants nodes.

Task #5: Implicit Opinions related to Verbnet verbs and roles

A human would easily understand that the people referred to by the sentence *People hope that the President will resign* have a rather negative opinion on the President because they envision his/her resignation. This simple sentence

²taken from *Sentiment Analysis and Opinion Mining*, Bing Liu, 2012

however lacks of terms explicitly indicating a positive or negative opinion, e.g. about the President, making it hard for a NLP-based tool to catch it. However, the term hope evokes a positive attitude towards what is referred to by the subordinate proposition the President will resign. This means that people refers to the holder of a positive opinion about a possible resign event (i.e., main topic) whose agent is the President (i.e. a subtopic). Intuitively, a subtopic is an entity that is indirectly targeted by an opinion sentence. In this case the opinion holder indirectly expresses an opinion on the President, while it directly expresses an opinion on a resign event. Being a resignation a generally negative event for its agent, a positive judgement of it implies a negative one on its agent. In this task a list of VerbNet verbs roles (around 20 verbs will be indicated by the challenge chairs) should be annotated and the developed algorithm should take into account the annotated resource in order to answer to sentences as the one mentioned at the beginning of this section. Challengers might have a look at a similar research paper³ where a resources called Sentilonet and a tool Sentilo have been developed for the same purpose. The expected output should be a polarity value (positive, negative) on detected VerbNet roles of identified verbs in the sentences included in the list of selected verbs. The output format for such a task is the following:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<Sentences>
  <sentence id="348:0">
    <text>Tom is happy that the President and the VicePresident were condemned.</text>
    <Sentilonet>
      <opinion targetverb="were condemned" verbrole="Theme"
        roleobject="President" polarity="negative"/>
      <opinion targetverb="were condemned" verbrole="Theme"
        roleobject="VicePresident" polarity="negative"/>
    </Sentilonet>
  </sentence>
</Sentences>
```

Input is the same without the Sentilonet tag and its descendants nodes.

³*Sentilo: Frame-Based Sentiment Analysis*, Diego Reforgiato Recupero, Valentina Pre-sutti, Sergio Consoli, Aldo Gangemi, Andrea Giovanni Nuzzolese, Cognitive Computation, 2014, Springer Science+Business Media New York 2014.

Judging and Prizes

One award will be given for each task (the winner of each task will be the one with the highest score in precision-recall analysis) and one more award will be given for the most innovative approach (the system with the best use of common-sense knowledge and semantics and innovative nature of approach).

The awards will consist in Springer vouchers and cash prize (depending on sponsors availability).

Besides, each challenge paper will be included in a Springer book as already done in challenge editions of 2014 and 2015.