



Institute of Informatics
Logics and Security Studies

ECO Annotation Guidelines version 1.2

Nick Webb, Sharon Small, Samira Shaikh
September 1, 2010

ILS Tech Report: 012

Available for download from:
<http://www.ils.albany.edu>

1 Introduction

In the ECO project, we are trying to automatically make judgments of opinion over concepts in text data taken from blogs on the Internet. Blogs are often opinionated pieces dealing with issues and topics effecting a community of people, a community to which the author of an individual blog belongs. We are training machine learning systems that can identify, in raw blog material, those concepts which are talked about either positively or negatively, depending on the target community.

To aid us in this task, we need annotators to identify words in blogs which are talked about positively or negatively, and issues in blogs which are dealt with sympathetically or unsympathetically. Although we will define these concepts in more detail, it is important to understand that we mean things that are talked about explicitly in the text, not what we *think* is going on in the mind of the author. If it's not clear in the text, it should NOT be annotated.

2 Source Data

We have deliberately chosen blog data from two distinct and opposed groups. Our data is public source, and is taken from a range of blogs dealing with Israeli and Palestinian issues. You will see the text of these blogs, minus any of the formatting you might see if you visited the blogs themselves, although it's often possible to work out what the headline of the blog entry is (and any value judgments appearing in these titles should also be annotated). This data will include some opinions about concepts such as people, places, countries and entities. We want you to identify these concepts in the text and indicate the value judgment that is being made.

3 Types of judgments

There are FOUR kinds of judgments, that can be made either directly or indirectly, and we give examples of each of these. In short, there are strong judgments (where the text is very clear about the value given to a particular concept, both positive and negative), and less clear judgments (called sympathetic and unsympathetic), where the text gives you the strong impression that a certain concept is being valued, without being able to pin down where

Hamas is an evil terrorist organization which does not document its actions...

Figure 1: **Negative** Direct judgment

in the text that happens. To complicate things a little further, we also introduce the idea of co-reference, where pronouns (words like ‘he’, ‘she’, ‘it’ and ‘they’) are used to refer to other concepts in the text. We call judgments featuring these words *indirect* judgments.

At the end of this section, we present the annotation tool we have developed for this task, and show how each of these judgments can be made using the tool.

3.1 Direct judgments

Direct judgments are the most common, are those where there are clear indicators of judgment in the text, and where it is equally clear what concept those judgments apply to, such as direct statements that ‘X is bad’ or ‘Y is good’. These should be clear and unambiguous, and fall into both **positive** and **negative** judgments.

3.1.1 **Negative** Direct Judgments

Negative direct judgments involve an explicit concept (such as a group or person) and a statement of opinion about that group, such as the example in Figure 1. In this example, ‘Hamas’ is the concept which is being judged, and that judgment is negative (by virtue of the fact that the author is calling Hamas an evil terrorist organization). In this case, we would want to see the concept ‘Hamas’ annotated, with a value of ‘negative’. We are not concerned as to whether or not the valuation judgment is correct (in your opinion), only in capturing the judgments made by the author or authors of the blog material.

The Goldstone report transforms highly moral Israeli operations into crimes.

Figure 2: Positive Direct Judgment

3.1.2 Positive Direct Judgments

In the example shown in Figure 2, there is a positive judgment made of a concept. In this example, the Israeli operations are regarded by the author as being ‘highly moral’. This translates into a direct judgment on those actions. We want to see ‘Israeli operations’ highlighted, and a value of ‘positive’ assigned.

Note in this example that there can be an inference about another concept in this example, ‘Goldstone report’. We can infer that a report which ‘transforms’ something ‘highly moral’ into ‘crimes’ is viewed, by the author of this blog, as a negative judgment. If you feel that this is a negative judgment, please mark it accordingly.

3.2 A Note on Unit Size

In our two examples so far, we have highlighted three concepts; a single word concept (‘ Hamas ’), and two multi-word concepts (‘ Goldstone report ’, ‘ Israeli operations ’). Capturing these concepts is very important. Ideally, these should be short word or phrases to which a direct judgment is being applied. There are some instances where this is not explicit in the text, and we refer to these as indirect judgments. Additionally, we do NOT want to capture common words like ‘ The ’, unless they make up a critical part of the annotation. For example, if the concept is ‘ The Palestinians... ’, we want to capture ‘ Palestinians ’ only. However, if the concept is ‘ President of the United States ’, then clearly we capture the entire phrase, including ‘ the ’.

3.3 Complex Noun Phrases

There are some noun phrases that seem inherently positive or negative. For example ‘Israeli terrorists’ can be viewed as negative or unsympathetic, whereas ‘Israeli martyrs’ is more positive or sympathetic (it depends on the actual text as to which of these you choose). However, there are other phrases such as ‘Christian children’ which you might think are positive (or negative), just because children are innocent. The solution for us is to turn one of the words in your mind into a verb, and see if it feels good or bad. For example, to ‘bomb’ someone seems bad. To ‘martyr’ seems better. To ‘children’ someone makes no sense. On those grounds, we should annotate the first two, and not the last one.

3.4 Indirect Judgments

Indirect judgments are those where the main entity that is being judged is not explicitly mentioned in the text, but is instead referred to somehow. Sometimes these references can be semi-explicit. In the example shown in Figure 3, ‘it’ is clearly the same as the ‘Goldstone report’ from the previous sentence. In this example, we want ‘report’ to be one of the concepts annotated, and given a judgment.

When annotating co-referent judgements, we would like the annotator to include the exactly correct reference, that is, which phrase (probably from the previous sentence) is being referred to. In example 3, that means including ‘Goldstone report’ as the referent. When selecting a co-referent judgment, the annotation tool should prompt the annotator to select the reference expression. You should select exactly the concept in the text that the expression refers to.

Note also that in this example, we see another type of coreference, where ‘them’ is used to refer to the ‘Israeli operations’ in the previous sentence. There is no clear direct judgment made on these ‘Israeli operations’ in this second sentence, and so we do not annotate them as such. However, the annotator might feel that the author is sympathetic to these ‘Israeli operations’, and can annotate them accordingly as described in Section 3.5.

The **Goldstone report** transforms highly moral **Israeli operations** into crimes. Unable to ignore the credibility of the reporting of these events, **it** twists them into a criminal attack on civilians.

Figure 3: Coreferent **Negative** Judgment

The **Goldstone commission** is fatally biased. **It** heavily relied on the testimony of terrorists from both the **Hamas** and **PA** sides.

Figure 4: **Unsympathetic** Judgment

The Goldstone commission is fatally biased against Israel.

Figure 5: Sympathetic Judgment

3.5 Sympathetic & Unsympathetic Judgments

Finally, we would like the annotators to capture what appear to be sympathetic or unsympathetic statements about concepts that occur in the blog. Taking the example shown in Figure 4, we can see that although there is no direct or indirect reference to the ‘it’ (which refers to ‘Goldstone commission’), we might infer that, although there is no direct judgment, the author is being unsympathetic toward the ‘Goldstone commission’. In this case, we highlight the ‘it’, mark it as unsympathetic and make the coreference link to ‘Goldstone commission’. Note we also capture ‘PA’ and ‘Hammas’ as negative as they are referred to as terrorists.

Taking the example in Figure 5, we can consider that this sentence is somehow sympathetic to the Israeli position. Fundamentally, if the annotator believes that some statement is meant to convey a sympathetic or unsympathetic judgment of some entity, then we capture that where possible, by highlighting the concept in question, and selecting either sympathetic or unsympathetic judgment. In the example in Figure 5, ‘Israel’ is highlighted in green, indicating a sympathetic judgment.

4 EAT: The ECO Annotation Tool

The ECO annotation tool has been designed to facilitate the annotation of blog material. You need to load your source data into the tool, and annotate using the buttons provided. Remember to save your work often!

On the main screen of the tool, there are FIVE buttons. Four of these (POSITIVE, NEGATIVE, SYMPATHETIC and UNSYMPATHETIC) match with the direct judgments we want you to make on concepts in the text. If you see a concept to which you want to attach a direct positive judgment, highlight the word using the mouse, and click on the POSITIVE button. The

eat.Login File Edit Help EAT - ECO Annotation Tool Thu Aug 19 1:20 PM Q

A Dark Horse in New York Monday, April 5 6:00 pm EST The race for a **Hillary Clinton**'s old Senate seat is at the forefront of observations from the bulk of political observers. **Sen. Kristen Gillibrand** was appointed to the seat and faces a tough opponent from former Tennessee **Rep. Steve Cohen**. Further, the winner of that battle then could face rough opposition from a Republican. But, virtual unknown and social entrepreneur Jonathan Tasini is also vying for the seat and he has a deep connection to **Israel** and unexpected views on the peace process -LRB- quick point - Tasini has virtually no shot. Recent polls put him at well under 10 percent support -RRB-. **Tasini** lived in Israel for years and has family that currently lives near the West Bank. Tasini even has family that died in Yom Kippur War. Even with that background, Tasini's views fall pretty far from the general views of AIPAC he has called the blockade on the Gaza Strip intolerable and unconscionable' and supports a two state solution with a connected **West Bank** and Gaza Strip. **Tasini** however, his position on the conflict is prominently displayed on his Web site. The foreign policy section of his views is dominated by Israel-related policy, which is the first issues seen by a reader. He even includes a link to a petition to end the blockade of Gaza. These positions have gained Tasini few votes in New York, but they have caught the attention of some international organizations, including the Muslimah Writers Alliance, which promotes writing careers for Muslim women, and is not exactly the organization you would expect to support a Jewish-American politician with Israeli heritage. Conversely, **Gillibrand**'s campaign Web site does not mention the conflict in any prominent way. Tasini is n't going to win and the Gaza issue assuredly will not dominate voters' concerns on election day. But, he has garnered attention to issues of concern and perhaps picked up a few votes along the way.

Positive Negative Sympathetic Unsympathetic Concept

Figure 6: ECO Annotation Tool

word will now have a GREEN background, to indicate that it is positive. In Figure 6, ‘Hillary Clinton’ has been highlighted as a positive direct concept.

It is the same mechanism for indicating direct NEGATIVE, SYMPATHETIC and UNSYMPATHETIC judgments.

You should notice that once you assign a direct judgment to a word or phrase, the actual word turns blue, and is also underlined. If you leave your mouse cursor over one of these words, you’ll see the concept and a unique number. This number serves TWO purposes.

First, if you want to REMOVE the judgment from a concept, if you make a mistake, then you need to know what this number is, in order to remove the correct judgment. Hover your mouse cursor over the concept you wish to change, in order to find the unique number. In Figure 6, the mouse is over the UNSYMPATHETIC concept ‘Tasini’, and you can see the relevant pop up indicating that the concept is identified as ‘Tasini(4)’. If you wish to remove this judgment, select the EDIT menu from the top left of the tool, and click ‘remove direct judgment’. You will then be asked to CHOOSE which judgment to remove, and in this case ‘Tasini(4)’ will be one of the options.

The second use of the unique identifier is for indirect judgments. If there is a concept you wish to annotate as an indirect judgment, you need to know the unique identifier of the concept you will link to. For example, in Figure 6, we want to link ‘his’ (from ‘his position on the conflict is predominantly...’) to ‘Tasini’ from several sentences earlier. To create this annotation, we need to know the unique identifier for the correct ‘Tasini’, which we know to be ‘Tasini(4)’.

We should highlight the concept to be annotated (in this case ‘his’) and RIGHT CLICK with the mouse. When we right click, four options appear; indirect positive, indirect negative, indirect sympathetic and indirect unsympathetic. Choose which of the four fits the concept under consideration, and pick the correct concept from the list that is presented (in this case, you would choose ‘Tasini(4)’).

Annotating an indirect concept like this requires that the concept has been annotated somehow by the user. Often, the concept you wish to link to will have been annotated already through some direct judgment. However, it is also possible to annotate a concept that has not yet been annotated in any way. If the concept being referred to has not been annotated, then you first need to capture it using the CONCEPT button on the main screen. This ensures that this concept will then appear in the list of potential indirect

references.