

INTERNE REPORTS DER DFG-FORSCHERGRUPPE 437
„TEXTTECHNOLOGISCHE INFORMATIONSMODELIERUNG“

Projekt A2 „Sekimo“

Serengeti – A brief Starting Guide



Nils Diewald
Fakultät für Linguistik und Literaturwissenschaft
Universität Bielefeld
nils.diewald@uni-bielefeld.de

11. Februar 2008

Inhaltsverzeichnis

1	Introduction	3
2	Requirements	3
3	Serengeti	3
3.1	Group Menu	4
3.2	Document Menu	4
3.3	Annotation Menu	4
3.4	File Menu	5
3.5	Text Frame	5
3.6	Edit Form	5
3.7	Relation Frame	6
4	Annotation Process	6
5	Consensus Mode	7
6	Short Cuts	9
7	Links	10
	Literatur	10

1 Introduction

The SERENGETI Annotator is a web based client-server-application used for annotating semantic relations in text documents. It was developed in the „Sekimo“ project (A2), part of the distributed research group “Text Technological Modelling of Information” funded by the German Research Foundation (DFG).

Offering a user-friendly online annotation framework, SERENGETI is suitable for central corpus storage and distributed annotations on single documents.

2 Requirements

Currently, the software supports Mozilla Firefox¹

3 Serengeti

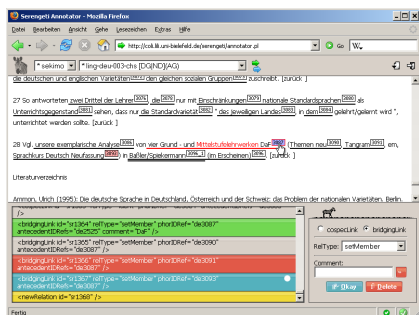


Before login, the graphical user interface of SERENGETI is separated into two parts: The login panel, where you have to enter your login data to start a new session, and the welcome screen, offering a short overview and links to several useful resources connected to SERENGETI.²

After you have logged in and have chosen your document to annotate (cf. *Document Menu*), the login panel changes its properties, now serving as application panel for managing and choosing groups, documents, and, contextual, annotations. On the right hand side of this panel, you get access to the *File Menu*. Here you are able to control your current annotation work. To exit your SERENGETI session, click on the Logout button next to it.

¹ SERENGETI is targeted at platform independence, so we have chosen Firefox (v 1.5 and up), which is freely available for Windows, Mac OS X and Linux systems, published by the Mozilla Project group under the terms of the Mozilla Public License at <http://www.mozilla.com/firefox/>. To start SERENGETI, just point your browser to: <http://coli.lili.uni-bielefeld.de/serengeti/>

² A demonstration account can be requested at the SERENGETI homepage.



The main frame displays the text of the document loaded, enriched by ID boxes to choose the fragments you want to be part of a relation.

Selection is done by clicking the respective boxes.³

On the bottom of the screen the *Relation Frame* appears with a list of all annotated relations on the left hand side and all tools

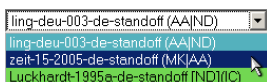
for editing them on the right hand side.

3.1 Group Menu



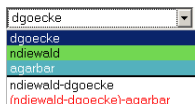
You can be member of more than one group, i. e. several annotation projects. Thus first you have to choose which annotation project you want to work in. After you have logged in, the first group in your list is automatically loaded, providing all documents of the group's folder in the *Document Menu*. Changing groups is possible at any time.

3.2 Document Menu



The documents of your annotation project are coloured according to their state of annotation: Documents with neutral (white) background colour have not been annotated by you up to now. Blue means you have started an annotation but haven't finished it yet. Green means you have finished annotating this document. In addition the state of work by other annotators is displayed in round (work in progress) or square brackets (finished).

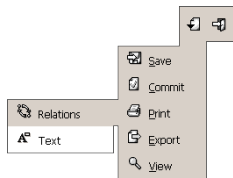
3.3 Annotation Menu



If you already annotated a document, the *Annotation Menu* displays your annotations of the chosen document. In case you loaded the document with special rights (cf. *Consensus Mode*), all annotations are shown by their annotator's names.

³ Markables, i. e. elements of the document that can take part in relations, are predefined

3.4 File Menu



The *File Menu* offers tools to operate on your annotation work. During the annotation process, saving is possible at any time by choosing *Save*. After completion you can declare your work as final version by choosing *Commit*. The *Print* option has a submenu, offering options to print either your relation list or the text. *Export* has a submenu, too, which differs between exporting your relation list as XML or CSV file or merging it with your input XML instance and exporting the complete version. The *View* submenu lists some features to ease your work: You can hide complete relations that are neither commented nor controversial, and you can visualize the markables that already take part in at least one relation.

3.5 Text Frame



The *Text Frame* shows the text to be annotated, formatted according to paragraphs, lists, tables and non-text sections. Additionally, markables are underlined and followed by boxes containing the markables' unique identifiers.⁴

3.6 Edit Form



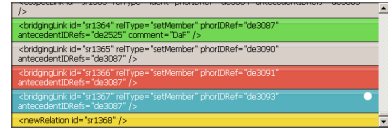
To create and edit relations, the *Edit Form* offers tools to select the features of the relation, and to add, delete or comment relations.

in the input XML.

⁴ Each identifier is realized as a clickable button, allowing to choose markables which take part in the annotation. By now, SERENGETI does not offer the possibility to add markables during the annotation process because this would mean to change the input data, which may result in inconsistent annotation data.

3.7 Relation Frame

All annotated relations are displayed in an XML representation in the *Relation Frame*. They can be chosen by clicking on the relations. The current relation is marked by a white dot at the end of the bar. With a double click the default relation is positioned right below the clicked relation.



All bars are highlighted differently according to their state:

<yellow />	The default relation.
<blue />	The relation is new, i. e. unsaved.
<grey />	The relation is old, i. e. saved.
<green />	The relation is commented.
<orange />	The relation is incomplete according to the annotation scheme.
<red />	The relation is deleted on the client side but not yet on the server side.
<purple />	The relation is controversial, i. e. made by only one annotator or there is an associated relation that differs in at least one aspect (cf. <i>Consensus Mode</i>).

4 Annotation Process

At the beginning, the relation list is empty except for one yellow coloured relation bar. This bar represents the default relation, marked with the white dot as current and ready to be edited. The next step is to choose the markables, that take place in this relation. By clicking the boxes after the markables in the *Text Frame*, they are highlighted and their IDs are added to the relation in the *Relation Frame*. The order depends on the type of semantic relations you are annotating, e. g. for working with anaphoric relations, the first markable is defined as the anaphoric element, the following markables as antecedents. By clicking the chosen markables again, they are deleted from the relation and deselected in the *Text Frame*.

Afterwards you should define the type of relation using the *Edit Form*. In case you are working with anaphoric relations and use the „Sekimo“

annotation scheme described in Stührenberg u. a. (2007) (which applies to the annotation in the demo group), you first have to decide between the main types *cospecLink* and *bridgingLink*. A new list automatically appears to define the subtype according to the main type, e.g. for *cospecLink* you can choose from *ident*, *propName*, *synonym*, *hyperonym*, *hyponym* etc. Additionally you can add comments to the selected relation.

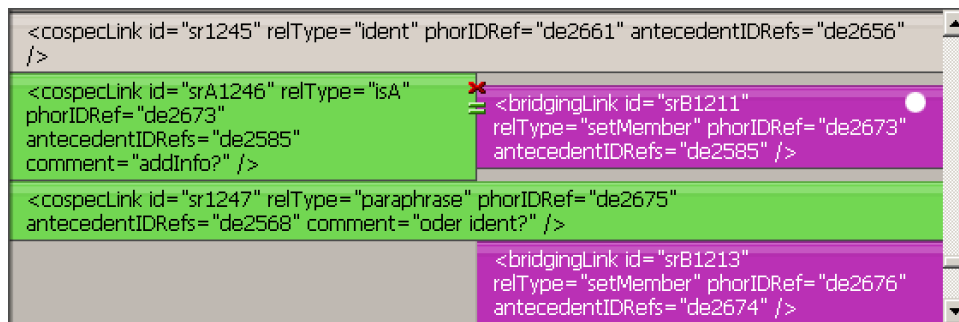
After completing the relation you can add it to the relation list by clicking *Okay*. In fact that means, you are adding a new default relation to the list and the former one is highlighted as a new annotated relation (cf. *Relation Frame*).

To edit these relations afterwards, simply choose them from the relation list and use the *Edit Form* and the markable boxes in the text to redefine the relation. The chosen relation is marked by the white dot at the right end of the bar. If you want to delete the relation completely, press the Delete button in the Edit Form; the relation will be highlighted red and will vanish after your next save. New relations can be placed anywhere in the relation list by simply double clicking on the bar in the list you want to precede the new one. The default relation bar then appears directly below the clicked relation. After having confirmed the relation, a new default relation appears at the end of the relation list.

5 Consensus Mode

Some users are allowed to view and edit all annotations in their group and compare multiple annotations on single documents. These users are called *Consensus Users*, allowing for verification of the quality of your annotation scheme by checking the inter-annotator-agreement and helps you achieving the best annotation results possible. *Consensus Users* have an extra blue arrow button in the application panel below the green one.























To start the *Consensus Mode*, you have to confirm the chosen document with the blue arrow button (instead of the green one). In the next step you can choose an annotation from the Annotation Menu. Again you have to decide whether to choose the green or the blue arrow. By confirming with the green arrow, you can edit the annotation. Changes are only saved in your folder, not in the annotation folder of the source annotation. By committing this version you declare an annotation as final. When clicking the blue arrow button, a new pulldown menu appears containing the rest of the



annotations. Here you can choose the annotation you want to compare with the previous one.⁵ After clicking the green arrow, both of the annotations are listed in the relation frame, sorted by their anaphoric element for example. Relations that appear in both annotations are highlighted grey, relations that appear only in one annotation are displayed on their respective sides. If relations share the same sort criteria value but differ in other aspects, they are displayed juxtaposed with border markers showing information on the type of difference. The annotation can be edited in the same way as single annotations, including deletion and confirmation of conflicting relations, as well as saving and committing. Conflictless compared annotations can be part of other comparisons again.

⁵ SERENGETI only allows to compare two annotations at one time.

6 Short Cuts

 + 	OKAY	Accept the relation definitions.
 + 	DELETE	Delete the current chosen relation.
 + 	SAVE	Save the current work to the server.
 + 	COMMIT	Declare the annotation as final.
 + 	PRINT RELATIONS	Print the list of relations with marked incomplete and commented relations. If containing juxtaposed relations, they will be printed out next to each other as well.
 + 	PRINT TEXT	Print out the formatted content of the Text Frame, including the underlined markables and the ID boxes.
 + 	EXPORT XML	Export the relations as a standalone XML file.
 + 	EXPORT CSV	Export the relations as a CSV file.
 + 	EXPORT ENTIRE XML	Export the relations as well as the input document, merged in a single XML file.
 + 	HIDEREADYREL	Show only relations that are incomplete, commented or controversial.
 + 	SHOWDISCOURSE ENTITIES	Show all discourse entities that have anaphoric meaning in at least one relation.

By now, the SHOWDISCOURSE ENTITIES functionality is available in the „Sekimo“ annotation scheme only.

7 Links

<http://www.text-technology.de/>

<http://www.text-technology.de/Sekimo/>

<http://www.coli.lili.uni-bielefeld.de/serengeti/>

<http://www.mozilla.com/firefox>

Literatur

[Stührenberg u. a. 2007] STÜHRENBERG, Maik ; GOECKE, Daniela ; DIEWALD, Nils ; CRAMER, Irene ; MEHLER, Alexander: Web-based Annotation of Anaphoric Relations and Lexical Chains. In: *The Linguistic Annotation Workshop (LAW), ACL 2007, Prague, 2007*