

KafAnnotator Manual

Version 0.1
August 2010

KYOTO Working paper: nr TR008, WP3 Capture

Knowledge-Yielding Ontologies for Transition-Based Organization
KYOTO (project nr. 211423)
FP7-ICT-2007-1



website: <http://www.kyoto-project.eu>

Authors:

Piek Vossen, Vrije Universiteit Amsterdam, p.vossen@let.vu.nl

Funded under project number 211423 in the 7th Frame work project in the area of Digital Libraries:
FP7-ICT-2007-1, Objective ICT-2007.4.2: Intelligent Content and Semantics

Table of Contents

1. Introduction.....	3
2. Getting started.....	5
2.1 Installation.....	5
2.2. Start the application.....	6
2.3. Quick overview.....	6
3. The tag file.....	15
4. The tag set file.....	15
5. The lexicon file.....	17
6. Annotating synsets.....	17
7. Converting annotations to training data.....	23
8. Converting annotations to event triplets.....	24

1. Introduction

KafAnnotator is a generic standalone program for the annotation of text represented in the Kyoto Annotation Format. The Kyoto Annotation Format (KAF) is a layered representation of natural language text that was developed in the Asian-European project KYOTO: <http://www.kyoto-project.eu>. In KAF different types of text analysis are stored as separate layers where each layer is anchored to a previous layer. Standard layers in KAF are: the token layer, the term layer and the chunk layer. An example of KAF is given below:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<text>
  <wf page="1" sent="1" wid="w1">The</wf>
  <wf page="1" sent="1" wid="w2">annual</wf>
  <wf page="1" sent="1" wid="w3">report</wf>
</text>
<terms>
  <term lemma="the" pos="D" tid="t1" type="close">
    <span>
      <target id="w1"/>
    </span>
  </term>
  <term lemma="annual_report" pos="N" tid="t2" type="open">
    <span>
      <target id="w2"/><target id="w3"/>
    </span>
  </term>
</terms>
<chunks>
  <chunk cid="c1" head="t2" phrase="NP">
    <span>
      <target tid="t1"/><target tid="t2"/>
    </span>
  </chunk>
</chunks>
```

The token layer (<text>) consists of a list of word tokens (<wf>) with identifiers in the linear order of

the text. The term layer is a list of word types (<term>) with separate identifiers for each term, where each term refers to one or more tokens (using the token identifier). For example, the term “annual_report” refers to two tokens. Chunks are constituents that consist of one or more terms. Chunks have their own identifiers and point to the identifiers of their terms and of the head term. More details on KAF can be found at:

http://www2.let.vu.nl/twiki/pub/Kyoto/TechnicalPapers/WP002_TR009_KAF_Framework.pdf

Annotations in KAF are stand-off annotations, i.e. any tags can be added in a new KAF layer with pointers to other elements in other layers. This offers large freedom in assigning tags to structures in text. The KafAnnotator makes use of this feature. When you read a KAF file, the tokens are listed in a column in the order of the original text together with their token identifiers. Any set of tags can then be assigned to any set of tokens (possibly disjoint) in the tag columns of the annotator. The tags are stored in a tag file that can be loaded with the original KAF file. In addition to tagging tokens, you can also tag word types (terms), constituents (terms) and sentences. Furthermore, it is possible to assign (different) tags to 3 levels. The same tokens can thus be tagged in 3 different ways.

Through the table interface it is possible to manipulate the text in many ways. You can sort all word forms or their lemmas in alphabetic order. You can also group them by part-of-speech, by tag and by tag identifier. You can at any point restore the order of the text by sorting on the order. In this way, you can choose to tag sequentially in text order or tag groups of tokens with the same form or properties in one tag action. The former way of tagging lets you focus on contextual sequences of words and structures, the latter way of tagging focuses on the types rather than the tokens and lets you focus on consistent and fast tagging of all tokens regardless of the context.

Another important feature of the KafAnnotator is that tags assigned to words are stored in a tag lexicon with the frequency of the tag for each word. Likewise, the annotator can assign the most frequent or dominant tag to any token. In some cases of tagging, for example, when assigning part-of-speech or word-senses, the most frequent tag applies in 80% to 90% of the cases. When the lexicon is sufficiently rich, you can assign the most frequent or dominant tag as the default tag and only change the tag when needed.

This manual is further structured as follows. In the next section, we give you a quick start on the annotator. In the subsequent sections, we explain the structures of the tag file, the tag set file and the lexicon file.

The KafAnnotator is licensed as open source software under license GNU General Public License version 3: <http://www.opensource.org/licenses/gpl-3.0.html>

2. Getting started

2.1 Installation

KafAnnotator requires Java JDK1.6 to be installed on a local machine. We assume that Java is installed and is added to the path settings of the machine. Ask your system manager for details.

To install the software unpack the archive file on any location. After unpacking the application, you will see the following directory structure:

```
kafannotator_v.0.1
  doc
    KafAnnotatorManual.pdf (this manual)
  examples
    11614.kaf
  images
  lib
    kaf.jar
    kafannotator.jar
  resources
    locations.xml
    tagSetLevel.txt
  kafannotator.sh
  kafannotator.bat
```

Path settings are relative to this directory structure. You should therefore leave the structure of the installation as it is. To remove the installation, you can simply delete the complete directory.

2.2. Start the application

The distribution includes a shell-script to launch the application from Linux/Unix and a bat file to launch the application on Windows.

Linux/Unix

Run the kafannotator.sh from the command line.

Windows

Double click the kafannotator.bat file or call it from the command line

2.3. Quick overview

After starting the application, you will see window with an empty table with columns and a menu bar. The columns are explained in the table below.

Column label	Explanation
Id	For storing the word token id
Word token	The word token read from the KAF file
Word type	The lemmatized word token or set of word tokens that makes up a word type
Pos	The part of speech of the word type as stored in the KAF file
Synset	Synset if the word type is present and if disambiguated
Tag1	First level tag for a word token
Tag 1 id	Identifier for the tagging event at level 1
Tag 2	Second level tag for a word token
Tag 2 id	Identifier for the tagging event at level 2
Tag 3	Third level tag for a word token
Tag 3 id	Identifier for the tagging event at level 3

Tag 4	Fourth level tag for a word token
Tag 4 id	Identifier for the tagging event at level 4
Tag 5	Fifth level tag for a word token
Tag 5 id	Identifier for the tagging event at level 5
Tag 6	Sixth level tag for a word token
Tag 6 id	Identifier for the tagging event at level 6
Order	Sequence number of the token
Status	Status of the word token

The first thing to do is to load a KAF file. Open the *File* menu and select the menu *Open KAF File*. An example of a KAF file is provided in the examples folder in the distribution. Select the file: examples/11767.mw.wsd.ne.kaf. The KAF file is loaded and the table is filled with the word tokens and their types, as shown in the next screen dump.

KAF Annotator Table View															Tag level 1 Tag level 2 Tag level 3 Tag level 4 Tag level 5 Tag level 6				
	Word token	Word type	Pos	Synset	Tag1	Tag1 id	Tag2	Tag2 id	Tag3	Tag3 id	Tag4	Tag4 id	Tag5	Tag5 id	Tag6	Tag6 id	Order	Status	Sentence
w1	A	a	D		0	0	0	0	0	0	0	0	0	0	0	0	false	1	
w2	Health	health	N		0	0	0	0	0	0	0	0	0	0	0	1	false	1	
w3	and	and	O		0	0	0	0	0	0	0	0	0	0	0	2	false	1	
w4	Restoration	restoration	N		0	0	0	0	0	0	0	0	0	0	0	3	false	1	
w5	Assessment	assessment	N		0	0	0	0	0	0	0	0	0	0	0	4	false	1	
w6	of	of	P		0	0	0	0	0	0	0	0	0	0	0	5	false	1	
w7	the	the	D		0	0	0	0	0	0	0	0	0	0	0	6	false	1	
w8	Chesapeake	chesapeake bay	N		0	0	0	0	0	0	0	0	0	0	0	7	false	1	
w9	Bay	chesapeake bay	N		0	0	0	0	0	0	0	0	0	0	0	8	false	1	
w10	and	and	O		0	0	0	0	0	0	0	0	0	0	0	9	false	1	
w11	Watershed	watershed	N		0	0	0	0	0	0	0	0	0	0	0	10	false	1	
w12	in	in	P		0	0	0	0	0	0	0	0	0	0	0	11	false	1	
w13	2008	2008	N		0	0	0	0	0	0	0	0	0	0	0	12	false	1	
w14	CBP/TRS	CBP/TRS	O		0	0	0	0	0	0	0	0	0	0	0	13	false	1	
w15	293-09	293-09	N		0	0	0	0	0	0	0	0	0	0	0	14	false	1	
w16	EPA-903-R-09-...	EPA-903-R-09-...	O		0	0	0	0	0	0	0	0	0	0	0	15	false	1	
w17	March	March	2009	N	0	0	0	0	0	0	0	0	0	0	0	16	false	1	
w18	2009	March, 2009	N		0	0	0	0	0	0	0	0	0	0	0	17	false	1	
w19	Bay	bay	G	eng-30-...	0	0	0	0	0	0	0	0	0	0	0	18	false	1	
w20	Barometer	barometer	N	eng-30-...	0	0	0	0	0	0	0	0	0	0	0	19	false	1	
w21	Chesapeake	Chesapeake	O		0	0	0	0	0	0	0	0	0	0	0	20	false	2	
w22	bay	bay	O	eng-30-...	0	0	0	0	0	0	0	0	0	0	0	21	false	2	
w23	Program	program	N		0	0	0	0	0	0	0	0	0	0	0	22	false	2	
w24	A	a	D		0	0	0	0	0	0	0	0	0	0	0	23	false	3	
w25	Watershed	watershed	N		0	0	0	0	0	0	0	0	0	0	0	24	false	3	
w26	Partnership	partnership	N		0	0	0	0	0	0	0	0	0	0	0	25	false	3	
w27	www.chesapeake...	www.chesapeake...	R		0	0	0	0	0	0	0	0	0	0	0	26	false	4	
w28	AbouttheChesAP...	AbouttheChesAP...	O		0	0	0	0	0	0	0	0	0	0	0	27	false	5	
w29	The	the	D		0	0	0	0	0	0	0	0	0	0	0	28	false	6	
w30	Chesapeake	chesapeake bay	N		0	0	0	0	0	0	0	0	0	0	0	29	false	6	
w31	Bay	chesapeake bay	N		0	0	0	0	0	0	0	0	0	0	0	30	false	6	
w32	is	be	V		0	0	0	0	0	0	0	0	0	0	0	31	false	6	
w33	an	an	D		0	0	0	0	0	0	0	0	0	0	0	32	false	6	
w34	estuary	estuary	N	eng-30-...	0	0	0	0	0	0	0	0	0	0	0	33	false	6	
w35					0	0	0	0	0	0	0	0	0	0	0	34	false	0	
w36	a	a	D		0	0	0	0	0	0	0	0	0	0	0	35	false	6	
w37	body	body	N		0	0	0	0	0	0	0	0	0	0	0	36	false	6	
w38	of	of	P		0	0	0	0	0	0	0	0	0	0	0	37	false	6	
w39	water	water	N		0	0	0	0	0	0	0	0	0	0	0	38	false	6	
w40	where	where	C		0	0	0	0	0	0	0	0	0	0	0	39	false	6	
w41	fresh	fresh	G		0	0	0	0	0	0	0	0	0	0	0	40	false	6	
w42	and	and	O		0	0	0	0	0	0	0	0	0	0	0	41	false	6	
w43	salt	salt	V		0	0	0	0	0	0	0	0	0	0	0	42	false	6	
w44	water	water	N		0	0	0	0	0	0	0	0	0	0	0	43	false	6	
w45	mix	mix	N		0	0	0	0	0	0	0	0	0	0	0	44	false	6	
w46					0	0	0	0	0	0	0	0	0	0	0	45	false	0	
w47	It	it	N	eng-30-...	0	0	0	0	0	0	0	0	0	0	0	46	false	7	
w48	is	be	V		0	0	0	0	0	0	0	0	0	0	0	47	false	7	
w49	the	the	D		0	0	0	0	0	0	0	0	0	0	0	48	false	7	
w50	largest	large	G		0	0	0	0	0	0	0	0	0	0	0	49	false	7	
w51	estuary	estuary	N	eng-30-...	0	0	0	0	0	0	0	0	0	0	0	50	false	7	

Messages: Loaded from:/Code/vu/KybotEvaluation/data/11767.mw.wsd.ne.kaf. Nr of wordtokens=16145

In the message field at the bottom of the screen, you will see the message:

Loaded from:/Code/vu/KybotEvaluation/data/11767.mw.wsd.ne.kaf. Nr of wordtokens=16145

It shows the path to the KAF file and the number of word tokens that are loaded.

The text can be read in the second column top down. The KAF identifiers for the word tokens are in the first column. These identifiers are used to connect the tag output to the original KAF file. You can click on the column headers to sort the table according to any column. Try sorting the table in different ways to see the effect for grouping word tokens. To restore the reading order of the words, use the Order column for sorting the words.

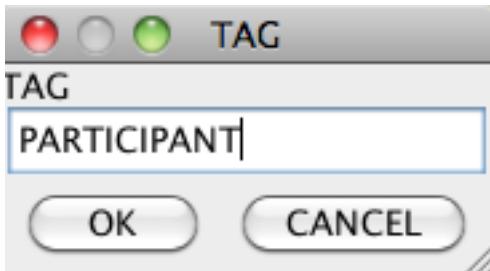
Only the Tag columns and the Status column can be edited. You can type in tags manually in the Tag

columns and tick and untick the status column immediately. For each work token you can specify at most 6 tags. However, it is better to use a predefined tag set to do the tagging. Tag sets can be created and modified through the interface or stored in any text file. Once a tag set file is created you can load the tag set at any time.

We start by creating tags for level 1. Open the *Tag level 1* menu. Below this you see the following submenus:

- Add tag
- Tag Tokens ▶
- Tag types ▶
- Tag constituent ▶
- Tag sentence ▶
- UNTAG
- Dominant tag

Tag tokens, types, constituent and sentence are used to assign tags from a tag set to any of these structures in the Tag1 column. However the first time you start, there are no submenus to select a tag. Therefore, we select the *Add tag* menu to add a tag. We fill in the tag PARTICIPANT and press OK.



We select the *Add tag* menu again to add the tags ROLE-DONE-BY, ROLE-PATIENT, LOCATION and TIME.

Next select some rows in the table. In this example, we select the rows for “in” and for “2008” and we will tag them as TIME. After selecting the rows, we open the *Tag level 1* menu again and open the submenu *Tag tokens*.

Tag level 1 at the word token level

KAF Annotation Tool									
File	Search	Confirm	Undo						
Tag level 1 Tag level 2 Tag level 3 Tag level 4 Tag level 5 Tag level 6									
Id	Word token	Word type	Pos	Tag1	Tag1 id	Tag2	Tag2 id	Tag3	Tag3 id
w1	A	D		0	0	0	0	0	0
w2	Health	N		0	0	0	0	0	0
w3	and	O		0	0	0	0	0	2
w4	Restoration	N		0	0	0	0	0	3
w5	Assessment	N		0	0	0	0	0	4
w6	of	P		0	0	0	0	0	5
w7	the	D		0	0	0	0	0	6
w8	Chesapeake	chesapeake bay	N	0	0	0	0	0	7
w9	Bay	chesapeake bay	N	0	0	0	0	0	8
w10	and	O		0	0	0	0	0	9
w11	Watershed	watershed	N	0	0	0	0	0	10
w12	in	P		0	0	0	0	0	11
w13	2008	2008	N	0	0	0	0	0	12
w14	CBP/TRS	CBP/TRS	O	0	0	0	0	0	13
w15	293-09	293-09	N	0	0	0	0	0	14
w16	EPA-903-R-09	EPA-903-R-09	O	0	0	0	0	0	15
w17	March	March 2009	N	0	0	0	0	0	16
w18	2009	March_2009	N	0	0	0	0	0	17
w19	Bay	bay	G	0	0	0	0	0	18
w20	Barometer	barometer	N	0	0	0	0	0	19
w21	Chesapeake	Chesapeake	O	0	0	0	0	0	20
w22	Bay	bay	G	0	0	0	0	0	21
w23	Program	program	N	0	0	0	0	0	22
w24	A	a	D	0	0	0	0	0	23
w25	Watershed	watershed	N	0	0	0	0	0	24
w26	Partnership	partnership	N	0	0	0	0	0	25
w27	www.chesapeak...	www.chesapeak...	R	0	0	0	0	0	26
w28	AbouttheChesAP...	AbouttheChesAP...	O	0	0	0	0	0	27
w29	The	the	D	0	0	0	0	0	28
w30	Chesapeake	chesapeake bay	N	0	0	0	0	0	29
w31	Bay	chesapeake bay	N	0	0	0	0	0	30
w32	is	be	V	0	0	0	0	0	31
w33	an	an	D	0	0	0	0	0	32
w34	estuary	estuary	N	0	0	0	0	0	33
w35	.			0	0	0	0	0	34
w36	a	a	D	0	0	0	0	0	35
w37	body	body	N	0	0	0	0	0	36
w38	of	of	P	0	0	0	0	0	37
w39	water	water	N	0	0	0	0	0	38
w40	where	where	C	0	0	0	0	0	39

Messages: Loaded tags from:/Code/vu/KafAnnotator/resources/KybotTagSet.txt. Nr of tags=18

w1 / March	march_zu09	N	U	U	U	U	U	U	U	16
w18 / 2009	March_2009	N	0	0	0	0	0	0	0	17

Tag level 1 at the word token level

We see that both “in” and “2008” have the tag *TIME* and share the same Tag1 tag id (column 6). A tag id is created for each tagging action that you perform. This way you can indicate that a sequence of tokens share a single tag. Using the submenus *Merge Tag Ids* and *Separate Tag Ids*, you can adapt the tag ids at any moment to joint them or make them unique for a selected set of tokens. Also note that the status of the rows is now ticked through the tagging action. You can also un-tick the status if are not sure about the status. The tags are kept but the status is not confirmed.

Just above “in 2008”, we see the tokens *Chesapeake* and *Bay* that belong to a single word type *chesapeake bay*. Select just the row for *Chesapeake* and select from the *Tag level 1* menu the *Tag types* submenu. Here you see the same set of tags from the tag set created. Select the *LOCATION* tag. If you selected the *Tag types* menu you will now see that both *Chesapeake* and *Bay* got the same tag. The tag has been

assigned to the word type level and thus is assigned to all word tokens belonging to the same type.

KAF Annotation Tool																	
File Search Confirm Undo			Tag level 1 Tag level 2 Tag level 3 Tag level 4 Tag level 5 Tag level 6 Order Status														
Id	Word token	Word type	Pos	Tag1	Tag1 id	Tag2	Tag2 id	Tag3	Tag3 id	Tag4	Tag4 id	Tag5	Tag5 id	Tag6	Tag6 id	Order	Status
w1	A	a	D		0		0		0		0		0		0	0	
w2	Health	health	N		0		0		0		0		0		0	1	
w3	and	and	O		0		0		0		0		0		0	2	
w4	Restoration	restoration	N		0		0		0		0		0		0	3	
w5	Assessment	assessment	N		0		0		0		0		0		0	4	
w6	of	of	P		0		0		0		0		0		0	5	
w7	the	the	D		0		0		0		0		0		0	6	
w8	Chesapeake	chesapeake bay	N	LOCATION	2		0		0		0		0		0	7	<input checked="" type="checkbox"/>
w9	Bay	chesapeake bay	N	LOCATION	2		0		0		0		0		0	8	<input checked="" type="checkbox"/>
w10	and	and	O		0		0		0		0		0		0	9	
w11	Watershed	watershed	N		0		0		0		0		0		0	10	
w12	in	in	P		0		0		0		0		0		0	11	
w13	2008	2008	N		0		0		0		0		0		0	12	
w14	CBP/TRS	CBP/TRS	O		0		0		0		0		0		0	13	
w15	293-09	293-09	N		0		0		0		0		0		0	14	
w16	EPA-903-R-09...	EPA-903-R-09...	O		0		0		0		0		0		0	15	
w17	March	March_2009	N		0		0		0		0		0		0	16	
w18	2009	March_2009	N		0		0		0		0		0		0	17	

Tag level 1 at the word type level

We can do the same for the constituent level. We scroll down to the rows *body of water* and select the row *body*. We do the same as before but now select the ROLE-PATIENT tag below the submenu *Tag constituent*. The result looks as follows:

KAF Annotation Tool																	
File Search Confirm Undo			Tag level 1 Tag level 2 Tag level 3 Tag level 4 Tag level 5 Tag level 6 Order Status														
Id	Word token	Word type	Pos	Tag1	Tag1 id	Tag2	Tag2 id	Tag3	Tag3 id	Tag4	Tag4 id	Tag5	Tag5 id	Tag6	Tag6 id	Order	Status
w29	The	the	D		0		0		0		0		0		0	28	
w30	Chesapeake	chesapeake bay	N		0		0		0		0		0		0	29	
w31	Bay	chesapeake bay	N		0		0		0		0		0		0	30	
w32	is	be	V		0		0		0		0		0		0	31	
w33	an	an	D		0		0		0		0		0		0	32	
w34	estuary	estuary	N		0		0		0		0		0		0	33	
w35	.	.			0		0		0		0		0		0	34	
w36	a	a	D		0		0		0		0		0		0	35	
w37	body	body	N	ROLE-PATIENT	3		0		0		0		0		0	36	<input checked="" type="checkbox"/>
w38	of	of	P	ROLE-PATIENT	3		0		0		0		0		0	37	<input checked="" type="checkbox"/>
w39	water	water	N	ROLE-PATIENT	3		0		0		0		0		0	38	<input checked="" type="checkbox"/>
w40	where	where	C		0		0		0		0		0		0	39	
w41	fresh	fresh	G		0		0		0		0		0		0	40	
w42	and	and	O		0		0		0		0		0		0	41	
w43	salt	salt	V		0		0		0		0		0		0	42	
w44	water	water	N		0		0		0		0		0		0	43	
w45	mix	mix	N		0		0		0		0		0		0	44	
w46	.	.			0		0		0		0		0		0	45	
w47	it	it	N		0		0		0		0		0		0	46	
w48	is	be	V		0		0		0		0		0		0	47	
w49	the	the	D		0		0		0		0		0		0	48	

Tag level 1 at the constituent level

We now see that three tokens that make up a single constituent all received the same tag and tag identifier.

In the same way, you can also assign a tag to a complete sentence. Select a row for a single word token and choose a tag from the submenu *Tag sentence*. You will see that all the tokens from the same sentence receive the same tag and identifier.

Make different selections of rows and assign tags to tokens, types, constituents and sentences. If you

want to remove tags, select the rows and choose the UNTAG menu.

So far, we added tags to the Tag1 column. In order to assign tags to the Tag2, Tag3 up to Tag6 columns, we need to follow the same procedure are above. The KafAnnotator keeps separate tag sets for each level. You are free to define the tags for each level and it is also possible to use the same tag set for different levels.

Instead of making the tag sets manually, we are going to load the tag set from a file. Any tag added to a tag menu is stored in a tag file (default file name is: *tagSet.txt* stored in the resources folder). Open the *File* menu and select the *Load TAG set* submenu. You are asked to select a file and the popup opens in the resources folder of the installation. In this folder, there is already a file stored with the name *KybotTagSet.txt*. Select the file. The message box at the bottom of the screen now says:

Load tags level 2 from:/Code/vu/KafAnnotator/releases/kafannotator_v.0.1/resources/KybotTagSet.txt. Nr of tags=18

If we go now to the *Tag Level 2*, *Tag Level 4* and *Tag Level 6* submenus, we see a sub-submenu for *EVENT-SCOPE*, if we go to the *Tag Level 3* and *Tag Level 5* submenus, we see the same tags as for *Tag Level 1*.

We now assign the tag *EVENT* to the sentence with *Chesapeake Bay*. The result is shown in the next screen dump. The previous tags have been overwritten by the new tag. Instead of assigning to a sentence, you can remove the tag and assign them to any selection of tokens. We scroll down and assign the third level tag *EVENT* to a series of tokens in the next sentence. The result looks as follows.

KAF Annotation Tool																
File	Search	Confirm	Undo		Tag level 1	Tag level 2	Tag level 3	Tag level 4	Tag level 5	Tag level 6						
w20	Barometer	barometer	N		0	0	0	0	0	0	0	0	0	19	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w21	Chesapeake	Chesapeake	O		0	0	0	0	0	0	0	0	0	20	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w22	Bay	bay	G		0	0	0	0	0	0	0	0	0	21	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w23	Program	program	N		0	0	0	0	0	0	0	0	0	22	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w24	A	a	D		0	0	0	0	0	0	0	0	0	23	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w25	Watershed	watershed	N		0	0	0	0	0	0	0	0	0	24	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w26	Partnership	partnership	N		0	0	0	0	0	0	0	0	0	25	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w27	www.chesapeak...	www.chesapeak...	R		0	0	0	0	0	0	0	0	0	26	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w28	AbouttheChesAP...	AbouttheChesAP...	O		0	0	0	0	0	0	0	0	0	27	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w29	The	the	D	EVENT	6	0	0	0	0	0	0	0	0	28	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w30	Chesapeake	chesapeake bay	N	EVENT	6	0	0	0	0	0	0	0	0	29	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w31	Bay	chesapeake bay	N	EVENT	6	0	0	0	0	0	0	0	0	30	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w32	is	be	V	EVENT	6	0	0	0	0	0	0	0	0	31	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w33	an	an	D	EVENT	6	0	0	0	0	0	0	0	0	32	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w34	estuary	estuary	N	EVENT	6	0	0	0	0	0	0	0	0	33	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w35	,			EVENT	6	0	0	0	0	0	0	0	0	34	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w36	a	a	D	EVENT	6	0	0	0	0	0	0	0	0	35	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w37	body	body	N	EVENT	6	0	0	0	0	0	0	0	0	36	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w38	of	of	P	EVENT	6	0	0	0	0	0	0	0	0	37	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w39	water	water	N	EVENT	6	0	0	0	0	0	0	0	0	38	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w40	where	where	C	EVENT	6	0	0	0	0	0	0	0	0	39	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w41	fresh	fresh	G	EVENT	6	0	0	0	0	0	0	0	0	40	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w42	and	and	O	EVENT	6	0	0	0	0	0	0	0	0	41	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w43	salt	salt	V	EVENT	6	0	0	0	0	0	0	0	0	42	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w44	water	water	N	EVENT	6	0	0	0	0	0	0	0	0	43	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w45	mix	mix	N	EVENT	6	0	0	0	0	0	0	0	0	44	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w46	.			EVENT	6	0	0	0	0	0	0	0	0	45	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w47	It	it	N		0	0	0	0	0	0	0	0	0	46	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w48	is	be	V		0	0	0	0	0	0	0	0	0	47	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w49	the	the	D		0	0	0	0	0	0	0	0	0	48	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w50	largest	large	G		0	0EVENT	7	0	0	0	0	0	0	49	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w51	estuary	estuary	N		0	0EVENT	7	0	0	0	0	0	0	50	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w52	in	in	P		0	0EVENT	7	0	0	0	0	0	0	51	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w53	the	the	D		0	0EVENT	7	0	0	0	0	0	0	52	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w54	United	United States	R		0	0EVENT	7	0	0	0	0	0	0	53	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w55	States	United States	R		0	0EVENT	7	0	0	0	0	0	0	54	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w56	and	and	O		0	0EVENT	7	0	0	0	0	0	0	55	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w57	the	the	D		0	0EVENT	7	0	0	0	0	0	0	56	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w58	third	third	N		0	0EVENT	7	0	0	0	0	0	0	57	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w59	largest	large	G		0	0EVENT	7	0	0	0	0	0	0	58	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w60	in	in	P		0	0EVENT	7	0	0	0	0	0	0	59	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w61	the	the	D		0	0EVENT	7	0	0	0	0	0	0	60	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w62	world	world	N		0	0EVENT	7	0	0	0	0	0	0	61	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
w63	.				0	0	0	0	0	0	0	0	0	62	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w64	The	the	D		0	0	0	0	0	0	0	0	0	63	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w65	Bay	bay	N		0	0	0	0	0	0	0	0	0	64	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w66	is	be	V		0	0	0	0	0	0	0	0	0	65	<input type="checkbox"/>	<input checked="" type="checkbox"/>
w67	about	about	P		0	0	0	0	0	0	0	0	0	66	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Messages:

Sentence level tagging

Tagging is not automatically saved, just as with a text editor. You need to save your tags using the save *Save tagging* submenu under the File menu. The tags are by default saved in the same location as the KAF file, where the extension “.kaf” is replaced by the extension “.tag”. When a KAF file is loaded, the program always looks for a tag file with the KAF file. If you want to use another TAG file you can use the load TAG file menu. If you want to save the tags in another location you can use the *Save tagging as* menu. The program always uses the most-recent location for saving the tags until the program is terminated.

Now we have saved the tags, we re-load the KAF files. You can see the original KAF file without the

tags. From the File menu, we select the submenu Load TAG file. We select the TAG file that we just saved and the tags re-appear.

When you use the *Save tagging* menu, the tags are saved in the tag file for the KAF but also the lexicon is saved to a lexicon file. Once the lexicon file is built, you can use the submenu *Dominant Tag* below the *Tag level 1* menu. To see the effect, select all rows in the table and select the *Dominant Tag* submenu. You can see now that tokens that were tagged before got a tag based on the most frequent tag in the lexicon.

Try to experiment with the various options to tag and save the results while you are working just as with a regular text editor. The tag file and the lexicon are updated every time you save. At any point, you can also choose to save the tags and the lexicon file to another file or location using the *Save as* submenu below *File*. Note that saving to another file, means that the location of the tag file and the lexicon are also changed to the new file and regular saves will be stored there.

Tip: try using the ALT key to operate the menus with the keyboard and use arrows, TAB and ENTER to navigate through the table. The space bar can be used to toggle the status after selecting the Status column.

When you tag more than one file, you probably work with fixed tagsets and lexicons. The following steps should be done to tag these files in a consistent way:

1. Load the tagset: File/Load TAG set
2. Load the lexicon file: File/Load lexicon file
3. Load the KAF file: File/Open KAF file
4. Load the TAG file: File/Load TAG file (your previous work if any)
5. Assign tags
6. Save tagging: File/Save tagging

Other menus:

File/Export tags to train format machine learning	exports the annotation to a simple list format that can be used for machine learning
Export tags to triplets	exports the annotation to triplets for evaluating text mining
Confirm	confirm or unconfirm any set of annotations
Search	search for words or tags in the table

Undo

undo the last tagging action and restores the previous values

3. The tag file

A tag file is a tab separated file containing all tokens that have been tagged on separate rows. Each row contains the full set of columns from the table:

w1615	The	the	D	ROLE-PATIENT	49	EVENT-SCOPE	52		0	0
0			0	1338	true w1616	Chesapeake	chesapeake bay	N	ROLE-PATIENT	
49			EVENT-SCOPE	52	0	0	0		0	1339
true										
w1617	Bay		chesapeake bay	N	ROLE-PATIENT	49	EVENT-SCOPE	52	0	0
0			0	0	1340	true w1618	and	and	O	ROLE-PATIENT
49			EVENT-SCOPE	52	0	0	0	0	0	1341
true										
w1619	its	its	D	ROLE-PATIENT	49	EVENT-SCOPE	52		0	0
0			0	1342	true w1620	tributaries tributary	N	ROLE-PATIENT	49	EVENT-
SCOPE	52		0	0		0	0	1343	true	

4. The tag set file

The tag set file is a list of tags prefixed with the tag level:

```

1;EVENT
1;ROLE-PATIENT
1;ROLE-DONE-BY
1;LOCATION
1;TIME
2;EVENT-SCOPE
3;EVENT
3;TIME
3;LOCATION
3;ROLE-DONE-BY
3;ROLE-PATIENT
4;EVENT-SCOPE

```

```
5;EVENT
5;TIME
5;LOCATION
5;ROLE-DONE-BY
5;ROLE-PATIENT
6;EVENT-SCOPE
```

You can manually create the file and load it for tagging.

It is possible to create a hierarchical tag menu. You can manually edit the tag set file and prefix intermediate levels using a prefix string separated with “#”. Below is an example of a tagset file used in KYOTO that differentiates between tags that apply to endurants (objects and substances) only and tags that apply to perdurants (processes and states). The tags for endurants have a further subdivision for tags related to motion events:

```
1;EVENT
1;endurant#patient
1;endurant#done-by
1;endurant#instrument
1;endurant#use-of
1;endurant#participant
1;endurant#has-state
1;endurant#motion#path-of
1;endurant#motion#destination-of
1;endurant#motion#source-of
1;endurant#generic-location
1;endurant#product-of
1;perdurant#simple-cause-of
1;perdurant#result-of
1;perdurant#state-of
1;perdurant#purpose-of
1;perdurant#goal-of
1;LOCATION
1;TIME
```

Any number of levels can be added to the tags. Note that only the leaf-items can be assigned to text. If you want to be able to assign intermediate levels as undifferentiated tags, you can repeat the

intermediate level at the lowest level, e.g.:

```
1;NAMED-ENTITY#NAMED-ENTITY
1;NAMED-ENTITY#LOCATION
1;NAMED-ENTITY#TIME
```

If you add a new tag manually using the interface, you can also specify the level. For example, typing *endurant#experiencer* when adding a tag will insert *experiencer* below the *endurant* submenu.

5. The lexicon file

The lexicon file stores the tags used for a token. It has the following format:

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<lexicon>
    <entry>
        <form>Humber</form>
        <>tagData><tag>LOCATION</tag> <freq>1</freq></tagData>
    </entry>
    <entry>
        <form>2006</form>
        <>tagData><tag>TIME</tag> <freq>1</freq></tagData>
    </entry>
    <entry>
        <form>tangible</form>
        <>tagData><tag>PARTICIPANT</tag> <freq>1</freq></tagData>
        <>tagData><tag>STATE</tag> <freq>6</freq></tagData>
    </entry>
</lexicon>
```

The is used to assign the dominant tag to a token and to collect statistics on the tagging.

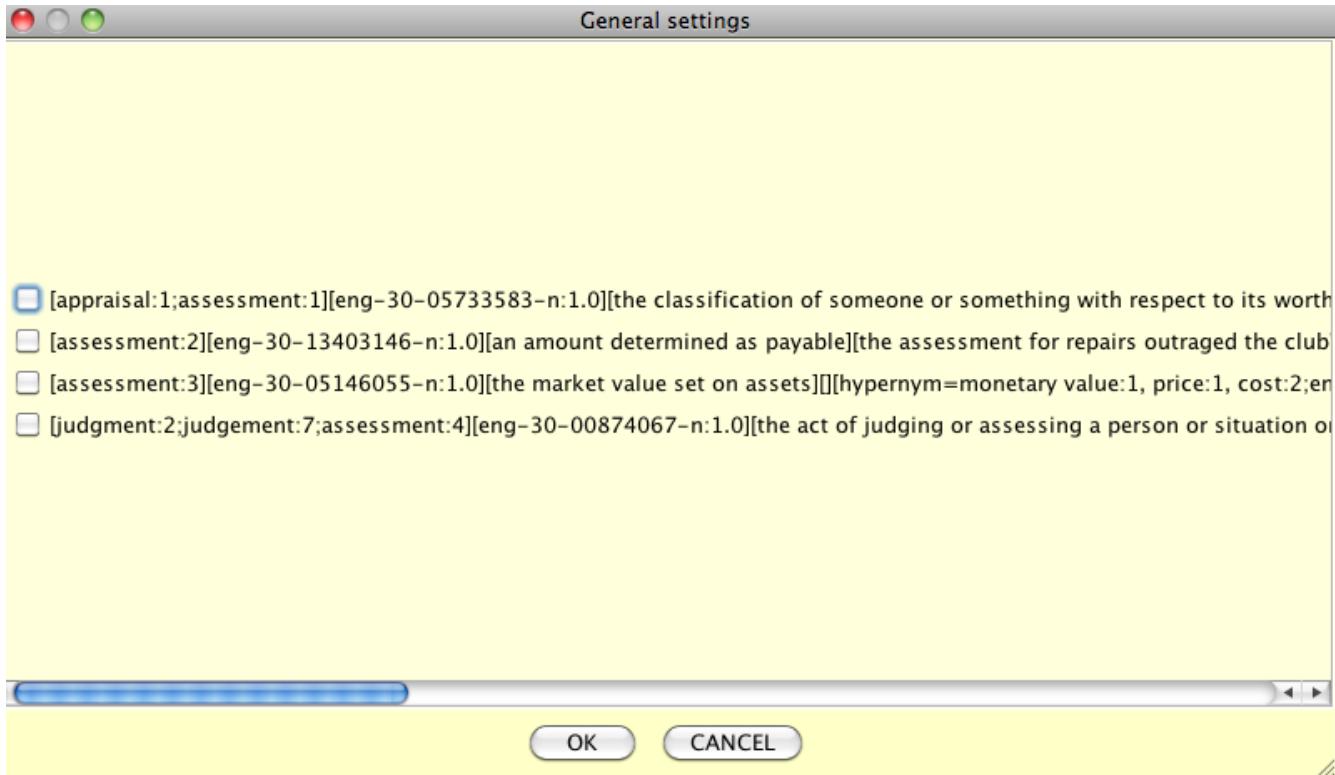
6. Annotating synsets

The synset column shows synset identifiers for word types that have a single synset associated with them in KAF. This can be the result of word-sense-disambiguation. It is possible to assign synsets to other word types or to choose another synset for word types with a single synset. To do that, you select a row with a word token and click with the right-mouse. As shown below, a pop-up menu with a list of wordnets from which you can select.. Selecting a menu will make a call to the online wordnet in the Kyoto project.

File	Search	Confirm	Undo		Tag level 1	Tag level 2	Tag level 3	T						
Id	Word token	Word type	Pos	Synset	Tag1	Tag1...	Tag2	Tag2...	Tag3	Tag3...	Tag4	Tag4...	Tag5	Ta
w1	A	a	D		0		0		0		0		0	
w2	Health	health	N		0		0		0		0		0	
w3	and	and	O		0		0		0		0		0	
w4	Restoration	restoration	N		0		0		0		0		0	
w5	Assessment	assessment	N		0		0		0		0		0	
w6	of	of	P		0		0		0		0		0	
w7	the	the	D		0		0		0		0		0	
w8	Chesapeake	chesapeake...	N		0		0		0		0		0	
w9	Bay	chesapeake...	N		0		0		0		0		0	
w10	and	and	O		0		0		0		0		0	
w11	Watershed	watershed	N		0		0		0		0		0	
w12	in	in		English wordnet	0		0		0		0		0	
w13	2008	200		Dutch wordnet	0		0		0		0		0	
w14	CBP/TRS	CBP		Spanish wordnet	0		0		0		0		0	
w15	293-09	293		Basque wordnet	0		0		0		0		0	
w16	EPA-903-R-...	EPA			0		0		0		0		0	
w17	March	March_2009	N		0		0		0		0		0	
w18	2009	March_2009	N		0		0		0		0		0	
w19	Bay	bay	G	eng-30-00395977-a:1.0	0		0		0		0		0	
w20	Barometer	barometer	N	eng-30-02794156-n:1.0	0		0		0		0		0	
w21	Chesapeake	Chesapeake	O		0		0		0		0		0	
w22	Bay	bay	G	eng-30-00395977-a:1.0	0		0		0		0		0	
w23	Program	program	N		0		0		0		0		0	
w24	A	a	D		0		0		0		0		0	
w25	Watershed	watershed	N		0		0		0		0		0	
w26	Partnership	partnership	N		0		0		0		0		0	
w27	www.chesape...	www.chesape...	R		0		0		0		0		0	
w28	AbouttheChe...	AbouttheChe...	O		0		0		0		0		0	
w29	The	the	D		0		0		0		0		0	
w30	Chesapeake	chesapeake...	N		0		0		0		0		0	
w31	Bay	chesapeake...	N		0		0		0		0		0	
w32	is	be	V		0		0		0		0		0	
w33	an	an	D		0		0		0		0		0	
w34	estuary	estuary	N	eng-30-09274500-n:1.0	0		0		0		0		0	

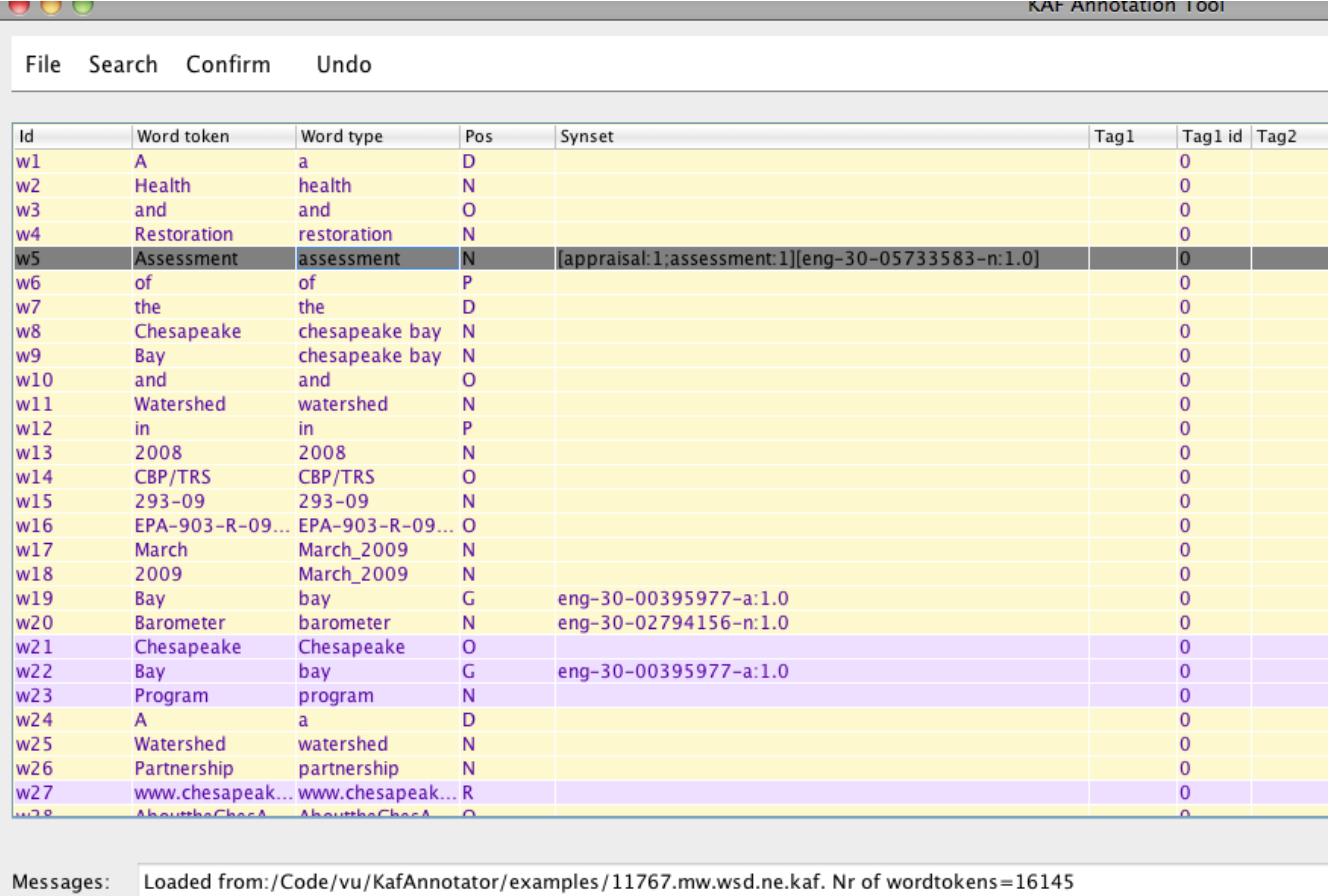
Messages: Loaded from:/Code/vu/KybotEvaluation/release/KybotEvaluation_v.0.1/data/11767.mw.wsd.ne.kaf. Nr of wordtokens=16145

The next window shows the different synsets associated with the senses of the word type. You can select a sense by the tick boxes and pressing OK.



The selected synset is inserted into the synset column as shown in the next screen dump. You can also tag single synsets that were selected. In some cases, the wrong part-of-speech is assigned and the wrong synset may be assigned as well. Making a call to the wordnet database, lets you correct this.

Note that the assigned senses are stored in the tag file and not in the original KAF file. When you load a KAF file, you need to load the tag file to see the assigned senses.



The screenshot shows the KAF Annotation Tool window. At the top, there's a menu bar with File, Search, Confirm, and Undo. Below the menu is a table with the following columns: Id, Word token, Word type, Pos, Synset, Tag1, Tag1 id, and Tag2. The table contains approximately 28 rows of data, each representing a word token with its corresponding type, part of speech, synset, and tags. The rows are color-coded by word type. A message at the bottom states: "Messages: Loaded from:/Code/vu/KafAnnotator/examples/11767.mw.wsd.ne.kaf. Nr of wordtokens=16145".

Id	Word token	Word type	Pos	Synset	Tag1	Tag1 id	Tag2
w1	A	a	D		0		
w2	Health	health	N		0		
w3	and	and	O		0		
w4	Restoration	restoration	N		0		
w5	Assessment	assessment	N	[appraisal:1;assessment:1][eng-30-05733583-n:1.0]	0		
w6	of	of	P		0		
w7	the	the	D		0		
w8	Chesapeake	chesapeake bay	N		0		
w9	Bay	chesapeake bay	N		0		
w10	and	and	O		0		
w11	Watershed	watershed	N		0		
w12	in	in	P		0		
w13	2008	2008	N		0		
w14	CBP/TRS	CBP/TRS	O		0		
w15	293-09	293-09	N		0		
w16	EPA-903-R-09...	EPA-903-R-09...	O		0		
w17	March	March_2009	N		0		
w18	2009	March_2009	N		0		
w19	Bay	bay	G	eng-30-00395977-a:1.0	0		
w20	Barometer	barometer	N	eng-30-02794156-n:1.0	0		
w21	Chesapeake	Chesapeake	O		0		
w22	Bay	bay	G	eng-30-00395977-a:1.0	0		
w23	Program	program	N		0		
w24	A	a	D		0		
w25	Watershed	watershed	N		0		
w26	Partnership	partnership	N		0		
w27	www.chesapeak...	www.chesapeak...	R		0		
w28	AbouttheChesA	AbouttheChesA	O		0		

It is possible to tag senses out of context. This is convenient if you have good reasons to believe that a word is used in the same meaning throughout the document. First you sort the text on the word type by clicking on the word type column. In the example shown below, you see now all occurrences of *waterway* out of context. You can select a synset for one occurrence as was explained above. Next, you select all rows for the same word type and do Select sense by right-clicking the mouse.

File	Search	Confirm	Undo							Tag
Id	Word token	Word type	Pos	Synset	Tag1	Tag1...	Tag2	Tag2 id	Tag	
w10479	way	way	N		0		0			
w11147	way	way	N		0		0			
w12764	way	way	N		0		0			
w13601	way	way	N		0		0			
w13661	ways	way	N		0		0			
w14069	way	way	N		0		0			
w15946	ways	way	N		0		0			
w16226	way	way	N		0		0			
w460	waterways	waterway	N		0		0			
w3888	waterways	water	Select sense	watercourse:3;waterway:2][en...	306	EVEN...	307			
w5198	waterways	waterway	N		0		0			
w10081	waterways	waterway	N		0		0			
w10991	waterways	waterway	N		0		0			
w12712	waterways	waterway	N		0		0			
w12809	waterways	waterway	N		0		0			
w15925	waterways	waterway	N		0		0			
w15976	waterways	waterway	N		0		0			
w16261	waterways	waterway	N		0		0			
w14251	Watershed	watershed ed...	N		0		0			
w14252	Education	watershed ed...	N		0		0			
w11	Watershed	watershed	N		0		0			
w25	Watershed	watershed	N		0		0			
w346	watershed	watershed	N		0		0			
w351	watershed	watershed	N		0		0			
w401	watershed	watershed	N	generi...	9	EVEN...	12			
w437	watershed	watershed	N		0		0			

Messages: Loaded from:/Code/vu/KybotEvaluation/release/KybotEvaluation_v.0.1/data/11767.mw.wsd.ne.3.k

The program will now take the first synset in the selection and apply it to all selected rows with the same word type. This is shown in the next screen dump. Note that you can next sort the text in the original order and re-evaluate each occurrence in context and where necessary correct the assignment that is made out-of-context.

File	Search	Confirm	Undo						
Id	Word token	Word type	Pos	Synset		Tag1	Tag1...	Tag2	Tag3
w10479	way	way	N			0		0	
w11147	way	way	N			0		0	
w12764	way	way	N			0		0	
w13601	way	way	N			0		0	
w13661	ways	way	N			0		0	
w14069	way	way	N			0		0	
w15946	ways	way	N			0		0	
w16226	way	way	N			0		0	
w460	waterways	waterway	N	[watercourse:3;waterway:2][en...		0		0	
w3888	waterways	waterway	N	[watercourse:3;waterway:2][en...	destin...	306	EVEN...	30	
w5198	waterways	waterway	N	[watercourse:3;waterway:2][en...		0		0	
w10081	waterways	waterway	N	[watercourse:3;waterway:2][en...		0		0	
w10991	waterways	waterway	N	[watercourse:3;waterway:2][en...		0		0	
w12712	waterways	waterway	N	[watercourse:3;waterway:2][en...		0		0	
w12809	waterways	waterway	N	[watercourse:3;waterway:2][en...		0		0	
w15925	waterways	waterway	N	[watercourse:3;waterway:2][en...		0		0	
w15976	waterways	waterway	N	[watercourse:3;waterway:2][en...		0		0	
w16261	waterways	waterway	N	[watercourse:3;waterway:2][en...		0		0	
w14251	Watershed	watershed ed...	N			0		0	
w14252	Education	watershed ed...	N			0		0	
w11	Watershed	watershed	N			0		0	
w25	Watershed	watershed	N			0		0	
w346	watershed	watershed	N			0		0	
w351	watershed	watershed	N			0		0	
w401	watershed	watershed	N		generi...	9	EVEN...	12	
w437	watershed	watershed	N			0		0	

Messages: Loaded from:/Code/vu/KybotEvaluation/release/KybotEvaluation_v.0.1/data/11767.mw.wsd

7 Converting annotations to training data

Many Machine Learning modules use a simple format for training data, where each token is presented on a separate line followed by the annotation separated by a tab:

the ROLE-PATIENT EVENT-SCOPE
overall ROLE-PATIENT EVENT-SCOPE
health ROLE-PATIENT EVENT-SCOPE
of ROLE-PATIENT EVENT-SCOPE
the ROLE-PATIENT EVENT-SCOPE
Chesapeake ROLE-PATIENT EVENT-SCOPE
Bay ROLE-PATIENT EVENT-SCOPE
did
not EVENT EVENT-SCOPE
improve EVENT EVENT-SCOPE
in TIME EVENT-SCOPE
2008 TIME EVENT-SCOPE
.

The ROLE-PATIENT EVENT-SCOPE ROLE-PATIENT EVENT-SCOPE ROLE-PATIENT EVENT-SCOPE
Bay ROLE-PATIENT EVENT-SCOPE ROLE-PATIENT EVENT-SCOPE ROLE-PATIENT EVENT-SCOPE
continues
to
have
poor EVENT EVENT-SCOPE
water EVENT EVENT-SCOPE
quality EVENT EVENT-SCOPE
,
degraded EVENT EVENT-SCOPE
habitats EVENT EVENT-SCOPE

and

low EVENT EVENT-SCOPE
 populations EVENT EVENT-SCOPE
 of EVENT EVENT-SCOPE
 many EVENT EVENT-SCOPE
 species EVENT EVENT-SCOPE
 of EVENT EVENT-SCOPE
 fish EVENT EVENT-SCOPE
 and EVENT EVENT-SCOPE
 shellfish EVENT EVENT-SCOPE

8. Converting annotations to event triplets

A special function is provided to convert KAF annotations to event triplet format: *File/Export to triplets*. The function reads the annotation tag file and generates a file with triplets. Such an annotation triplet file can be used as a gold standard for evaluating system output for event mining.

A triplet consists of:

- a relation
- a list of token ids that represent the event
- a list of token ids that represent a participant

Here is an example of a triplet:

```
<triplet id="16" relation="ROLE-PATIENT">
  <eventids>
    <event id="w12250"/>
  </eventids>
  <participantids>
    <participant id="w12253"/>
    <participant id="w12254"/>
    <participant id="w12255"/>
```

```
</participantids>  
</triplet>
```

If an event has multiple participants, a separate triplet is created for each event-participant pair. The triplet identifier is used to mark which triplets relate to the same event.

For the annotation of triplets the following requirements should be met:

1. At level 1, 3, and 5, one of the tags should be “EVENT”, other tags are free to choose (typical tags are roles of participants, time and location).
2. At level 2, 4, and 6, the tags should be “EVENT-SCOPE”

The annotators can tag any sequence of tokens using the tags at level 1, 3 and 5. In order to group annotations to the same event, they need to assign the EVENT-SCOPE tag to the set of tokens that belong to a single event. If participants (roles, time, location) belong to more than one event, they need to be annotated separately at another level (e.g. 3 or 5) and bound by separate EVENT-SCOPE. The next screen dump shows an example of tagging a KAF file in this way.

KAF Annotation Tool																Annotation layers						
Id	Word token	Word type	Pos	Tag1	Tag1 id	Tag2	Tag2 id	Tag3	Tag3 id	Tag4	Tag4 id	Tag5	Tag5 id	Tag6	Tag6 id	Order	Status	Tag level 1				
																		Tag level 5	Tag level 6			
w12212	been	be	V		0		0		0		0		0		0	0	11712	<input type="checkbox"/>				
w12213	included	include	V		0		0		0		0		0		0	0	11713	<input type="checkbox"/>				
w12214	into	into	P		0		0		0		0		0		0	0	11714	<input type="checkbox"/>				
w12215	an	an	D		0		0		0		0		0		0	0	11715	<input type="checkbox"/>				
w12216	ecosystem-based	ecosystem-based	O		0		0		0		0		0		0	0	11716	<input type="checkbox"/>				
w12217	fishery	fishery	N		0		0		0		0		0		0	0	11717	<input type="checkbox"/>				
w12218	management	management	N		0		0		0		0		0		0	0	11718	<input type="checkbox"/>				
w12219	plan	plan	N		0		0		0		0		0		0	0	11719	<input type="checkbox"/>				
w12220	.				0		0		0		0		0		0	0	11720	<input type="checkbox"/>				
w12221	During	during	P		0		0		0		0		0		0	0	11721	<input type="checkbox"/>				
w12222	2008	2008	N	TIME	1	EVENT-SCOPE	18	TIME	15	EVENT-SCOPE	24		0		0	0	11722	<input checked="" type="checkbox"/>				
w12223	,				0		0		0		0		0		0	0	11723	<input type="checkbox"/>				
w12224	biological	biological	G		0		0		0		0		0		0	0	11724	<input type="checkbox"/>				
w12225	briefs	briefs	N		0		0		0		0		0		0	0	11725	<input type="checkbox"/>				
w12226	and	and	O		0		0		0		0		0		0	0	11726	<input type="checkbox"/>				
w12227	background	background_infor...	N		0		0		0		0		0		0	0	11727	<input type="checkbox"/>				
w12228	information	background_infor...	N		0		0		0		0		0		0	0	11728	<input type="checkbox"/>				
w12229	for		P		0		0		0		0		0		0	0	11729	<input type="checkbox"/>				
w12230	such	such	C		0		0		0		0		0		0	0	11730	<input type="checkbox"/>				
w12231	plans	plan	N		0		0		0		0		0		0	0	11731	<input type="checkbox"/>				
w12232	were	be	V		0		0		0		0		0		0	0	11732	<input type="checkbox"/>				
w12233	completed	complete	V		0		0		0		0		0		0	0	11733	<input type="checkbox"/>				
w12234	.				0		0		0		0		0		0	0	11734	<input type="checkbox"/>				
w12235	Research	research	N		0		0	EVENT	11	EVENT-SCOPE	24		0		0	0	11735	<input checked="" type="checkbox"/>				
w12236	continued	continue	V		0		0		0		0		0		0	0	11736	<input type="checkbox"/>				
w12237	on	on	P		0		0		0		0		0		0	0	11737	<input type="checkbox"/>				
w12238	the	the	D		0		0		0		0		0		0	0	11738	<input type="checkbox"/>				
w12239	disease	disease	N	ROLE-DONE-BY	19			ROLE-PATIENT	23	EVENT-SCOPE	24		0		0	0	11739	<input checked="" type="checkbox"/>				
w12240	mycobacteriosis	mycobacteriosis	O	ROLE-DONE-BY	19	EVENT-SCOPE	18	ROLE-PATIENT	23	EVENT-SCOPE	24		0		0	0	11740	<input checked="" type="checkbox"/>				
w12241	.				0		0		0		0		0		0	0	11741	<input type="checkbox"/>				
w12242	Modeling	modeling	N		0		0		0		0		0		0	0	11742	<input type="checkbox"/>				
w12243	results	result	N		0		0		0		0		0		0	0	11743	<input type="checkbox"/>				
w12244	provided	provide	V		0		0		0		0		0		0	0	11744	<input type="checkbox"/>				
w12245	the	the	D		0		0		0		0		0		0	0	11745	<input type="checkbox"/>				
w12246	first	first	G		0		0		0		0		0		0	0	11746	<input type="checkbox"/>				
w12247	evidence	evidence	N		0		0		0		0		0		0	0	11747	<input type="checkbox"/>				
w12248	of	of	P		0		0		0		0		0		0	0	11748	<input type="checkbox"/>				
w12249	mycobacteriosis-a...	mycobacteriosis-a...	O		0		0		0		0		0		0	0	11749	<input type="checkbox"/>				
w12250	mortality	mortality	N	EVENT	2	EVENT-SCOPE	18		0		0		0		0	0	11750	<input checked="" type="checkbox"/>				
w12251	in	in	P		0		0		0		0		0		0	0	11751	<input type="checkbox"/>				
w12252	the	the	D		0		0		0		0		0		0	0	11752	<input type="checkbox"/>				
w12253	striped	striped	G	ROLE-PATIENT	22		0		0		0		0		0	0	11753	<input checked="" type="checkbox"/>				
w12254	bass	bass	G	ROLE-PATIENT	22		0		0		0		0		0	0	11754	<input checked="" type="checkbox"/>				
w12255	population	population	N	ROLE-PATIENT	22	EVENT-SCOPE	18		0		0		0		0	0	11755	<input checked="" type="checkbox"/>				
w12256	in	in	P		0		0		0		0		0		0	0	11756	<input type="checkbox"/>				
w12257	the	the	D		0		0		0		0		0		0	0	11757	<input type="checkbox"/>				
w12258	Bay	bay	N	LOCATION	14	EVENT-SCOPE	18		0		0		0		0	0	11758	<input checked="" type="checkbox"/>				
w12259	.				0		0		0		0		0		0	0	11759	<input type="checkbox"/>				
w12260	Scientists	scientist	N		0		0		0		0		0		0	0	11760	<input type="checkbox"/>				
w12261	also	also	A		0		0		0		0		0		0	0	11761	<input type="checkbox"/>	<input checked="" type="checkbox"/>			

Messages:

Using the annotation layers to annotate triplets for KYOTO evaluation

In this screen dump, disease mycobacteriosis is playing two different roles in two events.

Note that the tags EVENT and EVEN-SCOPE are obligatory to be able to derive triplets. The other tags are optional and any tag can be chosen.