

U-Compare Workbench

1. BASIC INFORMATION

Tool name

U-Compare Workbench

Overview and purpose of the tool

The U-Compare Workbench (Kano et al., 2009; Kano et al., 2011) is a graphical user interface that operates on top of the U-Compare platform. The U-Compare platform allows users to build and evaluate NLP workflows. Workflows consist of one or more components, consisting of corpus readers and tools, such as tokenisers, POS taggers, named entity recognisers, etc. Workflows can be built using any components that are compliant with the UIMA framework¹ (Ferrucci et al., 2006). The Workbench provides several facilities, including the following:

- Rapid construction of workflows by dragging and dropping components from a library onto the workflow canvas
- Graphical display of comparison of the performance of alternative workflows and evaluation against gold standard data
- Import of new UIMA components into the library
- Export of components/workflows

The core library of components provided with U-Compare includes several different types of tools, including sentence splitters, tokenisers, part-of-speech taggers, lemmatisers, named entity recognisers, etc. Currently, the majority of these are for English, with a focus on biomedical text. However, several UIMA components are currently under development for the processing of other European languages, such as Portuguese, Maltese, Romanian, Spanish, Catalan, Basque and Galician and French (Ananiadou et al., 2011; Thompson et al., 2011). These components will be added to the U-Compare library in the near future.

¹ <http://uima.apache.org/>

A short description of the algorithm

There is no algorithm to describe as such, as this is tool is a graphical user interface for the U-Compare platform.

2. TECHNICAL INFORMATION

Software dependencies and system requirements

The U-Compare workbench can be used in any environment in which Java 6 is available. At least the first time the system is run, an internet connection is required, since the most up-to-date relevant files are downloaded from the internet.

Installation

No specific installation is required. U-Compare can be started directly from the Internet by clicking on the “Start U-Compare” button on this page:

<http://www.nactem.ac.uk/ucompare/index.html>

However, it is preferable to start U-Compare from the command line, by downloading the file `UCLoader.class` from <http://u-compare.org/downloads/UCLoader.class>.

See also <http://www.nactem.ac.uk/ucompare/launch.html> for more information

Execution instructions

From the command line, the U-Compare workbench is started by running the `UCLoader.class` file, e.g.

```
java -Xms700m -Xmx1000m UCLoader
```

The `-Xms` and `-Xmx` specified the minimum and maximum memory allocated to U-Compare. The more memory is allocated, the quicker U-Compare will run. Note that the first time U-Compare is launched, relevant files will be downloaded from the internet. Therefore, the first time the system is launched, it may take a considerable amount of time to start up.

Input/Output data formats

Input data formats

The first component in a workflow must read in some input data. Currently, this can only be text (annotated or not), although other modalities, such as speech, are planned. This first component must be a “collection reader”, which reads the data to be processed into the UIMA Common Analysis Structure (CAS). This is the common data structure that can be accessed by all components in a workflow. Components obtain their input by reading annotations from the CAS, while the output of components is written to the CAS by creating new annotations, or updating existing annotations.

The library of U-Compare components includes several generic collection readers to read in plain text, e.g., from an input window or from a directory of files. Several corpus-specific readers (currently mainly are also provided to read annotated texts into the CAS. Collection readers for different annotated corpora can be added as required.

Output data format

As mentioned above, the output of a workflow is a set of annotations (possibly of various different types) that are added to the CAS. Different types of annotation viewers provided in the U-Compare workbench allow annotations to be viewed in different ways e.g., as simple text spans or as tree/HPSG structures (in the case of the display of parser results). It is possible for the annotations in the CAS to be written to an output file using a CAS consumer component. Components are provided in the U-Compare library to produce different types of output files, such as XMI, inline XML annotations, etc.

Integration with external tools

The workbench is intended to be run as a standalone application. However, workflows created using the workbench can be embedded into other applications.

3. CONTENT INFORMATION

In this section, some screenshots are provided to illustrate the functionality of the U-Compare Workbench. Further information about using U-Compare can be obtained from the documentation pages on the website:

- User manual: <http://www.nactem.ac.uk/ucompare/userguide/index.html>
- Developer manual: <http://www.nactem.ac.uk/ucompare/developerguide/index.html>

Figure 1 shows the main window of the U-Compare Workbench. On the right is the library of components, while on the left is the workflow canvas. To create a new workflow, components are simply dragged from the library onto the canvas, in the order in which they are to be executed. Components can also be reordered once they have been placed on the workflow canvas.

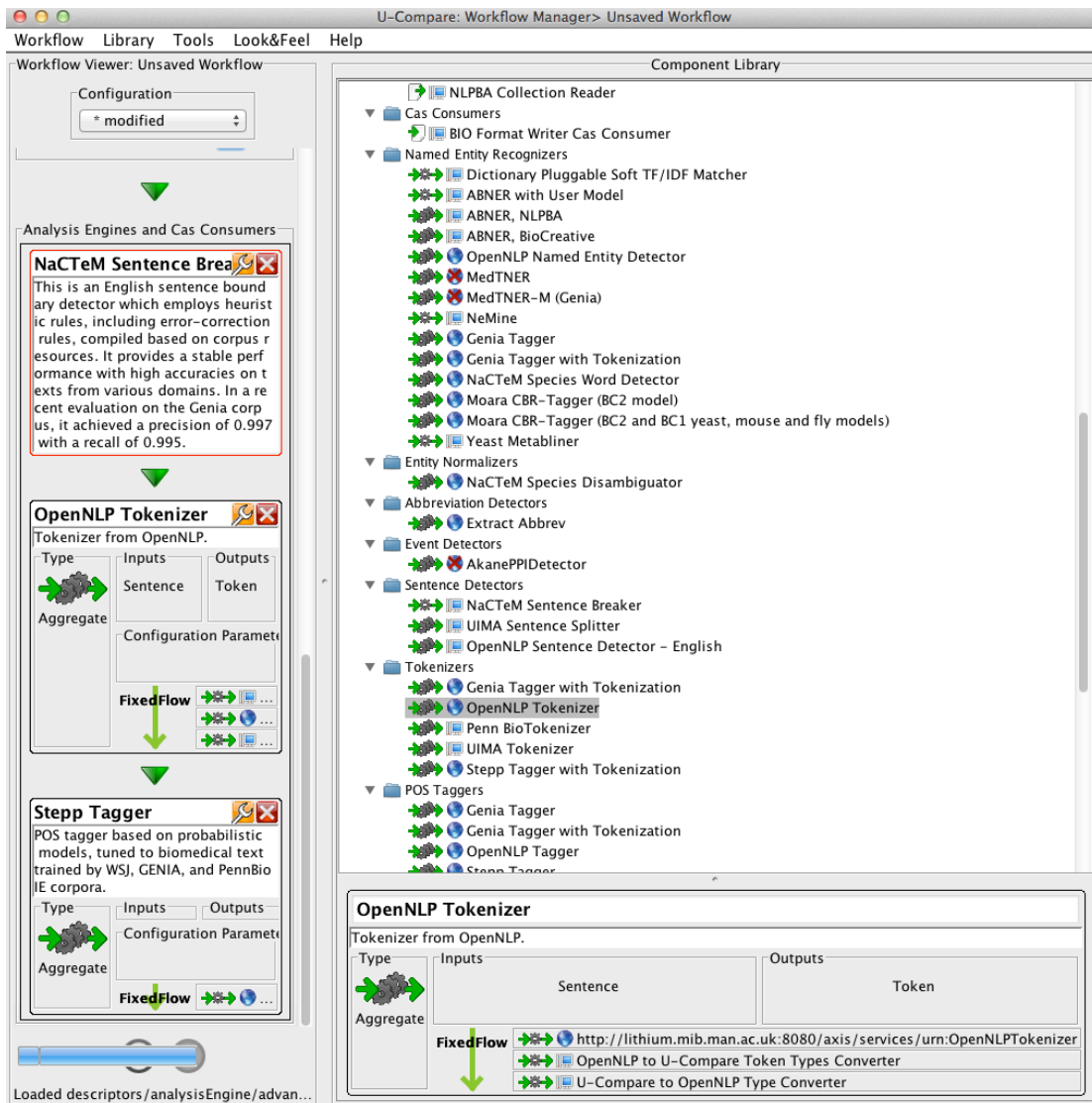


Figure 1: The main window of the U-Compare workbench

Figure 2 shows the display of the annotations in U-Compare's default annotation viewer, produced by the workflow shown in Figure 1. There are three types of annotations, highlighted using different colours, i.e. sentence annotations, tokens and part-of-speech annotations. Attributes associated with the different types of annotations can be viewed in a tabular format. This is illustrated on the right-hand side of Figure 2, which shows the table of attributes associated with the part-of-speech annotations. Each annotation stores the start and end offsets of the annotation, plus the part-of-speech, in the "posString" attribute. Clicking over a row in the table causes the corresponding annotation to be highlighted.

begin	end	posStr
0	3	DT
4	12	NNP
13	19	NNP
20	25	NNP
26	29	CC
30	35	NNP
36	45	NNP
46	54	IN
55	59	NN
60	62	IN
63	66	DT
67	75	NNP
76	86	NNPS
87	89	IN
90	97	NNP
98	101	VBZ
102	109	VBN
110	113	CD
114	117	NN
118	120	IN
121	122	DT
123	128	CD
129	132	NN
133	146	JJ
147	155	JJ
156	161	NN
162	170	VBC
171	181	NNS
182	185	IN
186	189	DT
190	203	NN
204	211	NN

The National Heart, Lung, and Blood Institute (NHLBI), part of the National Institutes of Health, has stopped one arm of a three arm multi-center, clinical trial studying treatments for the lung-scarring disease idiopathic pulmonary fibrosis (IPF) for safety concerns. The trial found that people with IPF receiving a currently used triple-drug therapy consisting of prednisone, azathioprine, and N-acetylcysteine (NAC) had worse outcomes than those who received placebos or inactive substances.

"These findings underscore why treatments must be evaluated in a rigorous manner," said Susan B. Shurin, M.D., acting director of the NHLBI. "This combination therapy is widely used in patients with IPF, but has not previously been studied in direct comparison to a placebo for all three drugs."

Figure 2: Annotation viewer

Figure 3 shows the output of a comparison workflow. The outputs of 2 named entity recognisers (ABNER-NLPBA and ABNER-BioCreative) are compared against a gold standard corpus (Aimed).

The system produces pairwise comparisons of the annotations, with different resources being assumed as the gold standard. Since Aimed is the gold standard corpus, only the rows in which Aimed is in the “Assumed Gold Standard” are meaningful in this comparison. In each row, several pieces of information are shown: the number of relevant annotations in the gold standard corpus (G), the number of annotations produced by the relevant tool (T), the number of matching annotations (M), the F1 score, precision (PR) and recall (RC). These figures are shown both for the collection as a whole, and for the individual documents in the corpus.

Assumed Gold Standard	Comparison Components	Total (All Documents)						pubmed abstract 11780382.xmi																	
		Boundary Match						Boundary Match																	
▼ .Protein	▲ .Protein	+	G	+	T	+	M	+	F1	+	PR	+	RC	+	G	+	T	+	M	+	F1	+	PR	+	RC
✓ Aimed	✓ ABNER-NLPBA	15	23	15	15	78.95	65.22	100.0	15	23	15	15	78.95	65.22	100.0										
✓ ABNER-NLPBA	✓ Aimed	23	15	15	15	78.95	100.0	65.22	23	15	15	15	78.95	100.0	65.22										
✓ Aimed	✓ ABNER-BioCreative	15	21	15	15	83.33	71.43	100.0	15	21	15	15	83.33	71.43	100.0										
✓ ABNER-BioCreative	✓ Aimed	21	15	15	15	83.33	100.0	71.43	21	15	15	15	83.33	100.0	71.43										

Figure 3: Output of comparison workflow

The comparison workflow is created using the following steps:

- 1) The Aimed corpus reader, which reads the gold-standard corpus, is dragged into the collection reader slot of the workflow canvas.

2) Since the ABNER-NLPBA and ABNER-BioCreative named entity recognition components both require Sentence annotations as input, a sentence splitter is first dragged onto the “Analysis Engines and CAS Consumers” section of the workflow canvas.

3) Next, a “Parallel Aggregate” component is added to the workflow, which will allow the outputs of the two named entity recognisers to be compared. Adding this component will cause an “Evaluation Iterator” component to be added to the workflow automatically. The workflow should now look similar to the one illustrated in Figure 4.

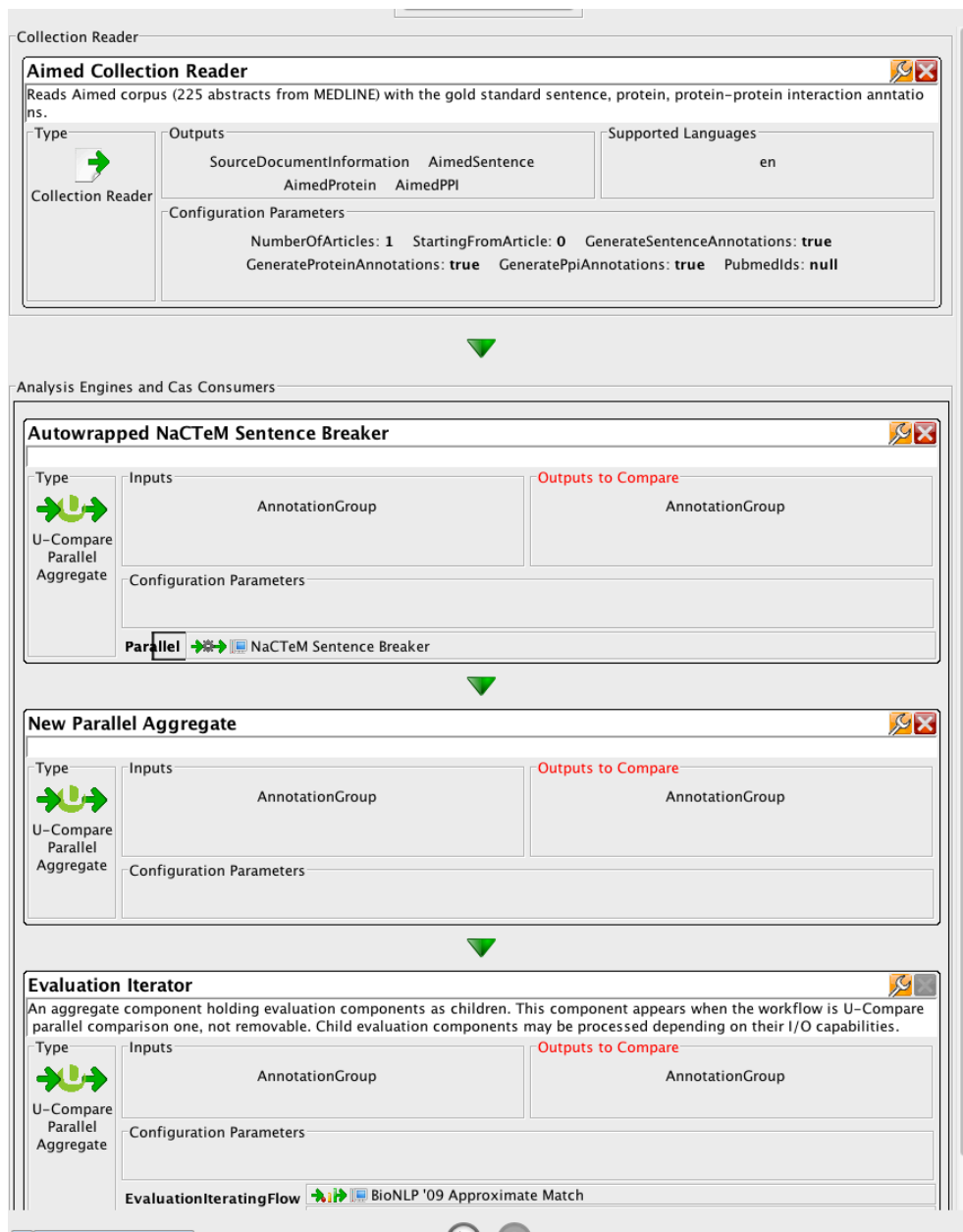



Figure 4: Comparison Workflow

4) Next, the components to be compared must be added to the parallel aggregate component. This is done by clicking on the  icon of the component to configure it. The configuration screen for this component includes a “canvas” onto which the components to be compared can be dragged and dropped. Thus, for the purposes of the current example, the “ABNER, NLPBA” and “ABNER, BioCreative” components are dragged onto this canvas.

5) Also on the configuration screen of the parallel aggregate component, it must be specified which type(s) of annotations are to be compared. This is done in the section headed “Outputs to Compare”. Within this section, the button labelled “Add output type” should be clicked. This will cause a type system viewer window to be displayed, with a graphical representation of the U-Compare type system. In this case, we wish to compare the “Protein” annotations that are produced by the two tools, and so the “Protein” type should be located in the tree diagram, and clicked upon. A part of the graphical tree representation of the type system is shown in Figure 5.

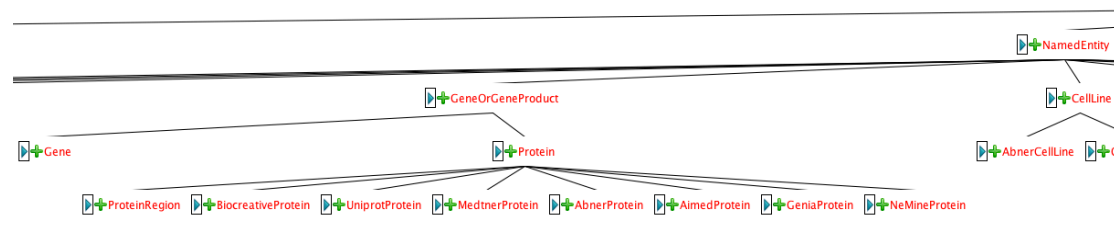


Figure 5: The graphical type system viewer

6) Once the above configuration steps have been completed, the configuration screen for the parallel aggregate component should appear as shown in Figure 6. The “Confirm Changes” button at the top of the screen can then be clicked, and the workflow can be run.

An evaluation table similar to the one shown in Figure 3 should then be displayed to compare the outputs of the 2 tools against the gold standard annotated corpus.

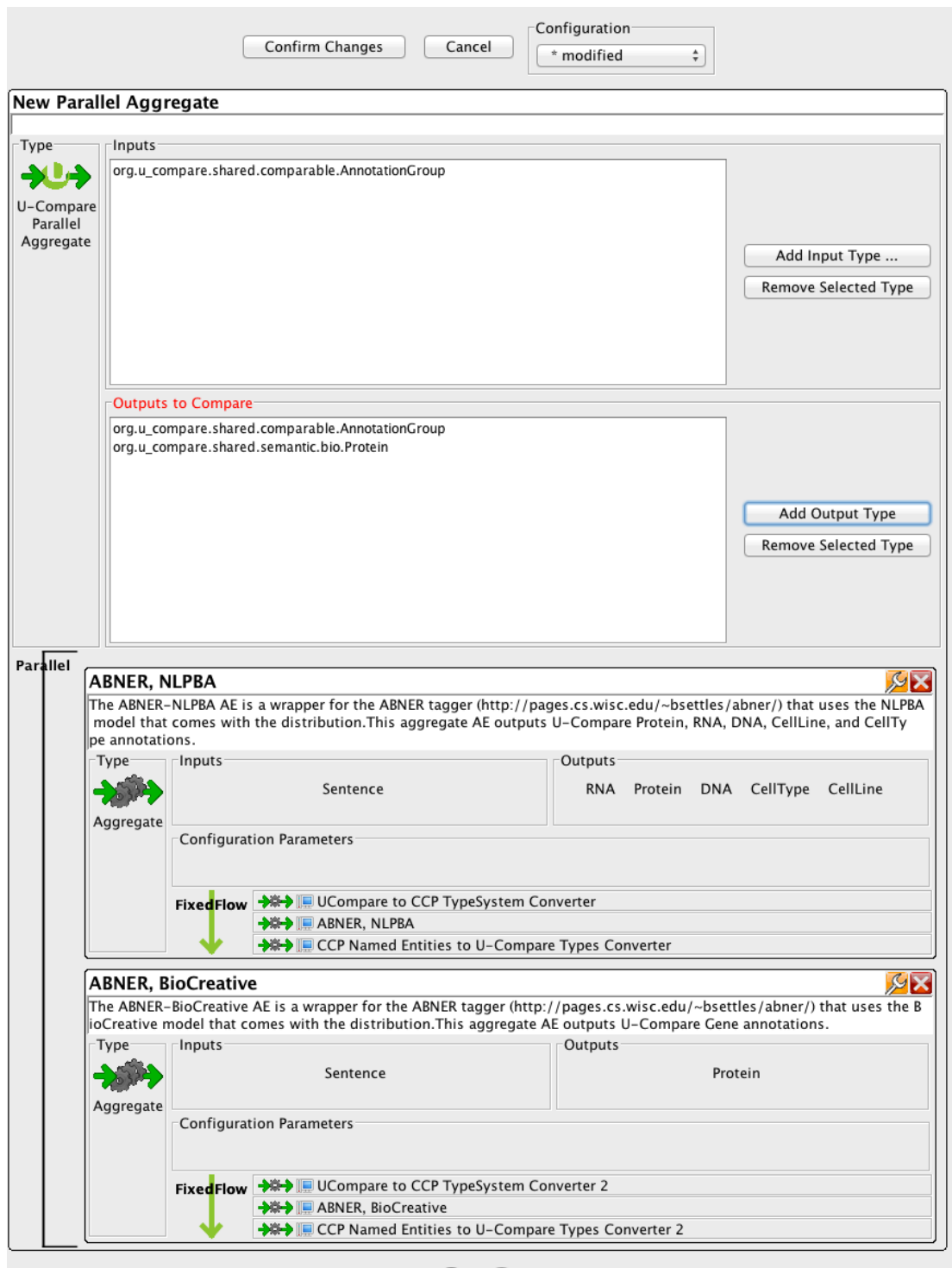


Figure 6: Completed configuration of parallel aggregate component

4. LICENCES

The U-Compare workbench is released under a dual license, the LGPL open source license (<http://www.gnu.org/licenses/lgpl.html>) or a commercial license. Please use the contact details below if you are interested in obtaining a commercial licence.

5. ADMINISTRATIVE INFORMATION

Contact

For further information, please contact Sophia Ananiadou:

sophia.ananiadou@manchester.ac.uk

6. REFERENCES

Ananiadou, S., Thompson, P., Kano, Y., McNaught, J., Attwood, T. K., Day, P. J. R., Keane, J., Jackson, D. and Pettifer, S.. (2011). "Towards Interoperability of European Language Resources". *Ariadne*, 67

Ferrucci, D., Lally, A., Gruhl, D., Epstein, E., Schor, M., Murdock, J. W., Frenkiel, A., Brown, E.W. , Hampp T., Doganata, Y., Welty, C., Amini, L., Kofman, G., Kozakov, L. and Mass, Y. (2006). Towards an Interoperability Standard for Text and Multi-Modal Analytics. IBM Research Report RC24122.

Kano, Y., Baumgartner, W. A., Jr., McCrohon, L., Ananiadou, S., Cohen, K. B., Hunter, L. (2009). "U-Compare: share and compare text mining tools with UIMA". *Bioinformatics*, vol. 25, no. 15, 1997-1998.

Kano, Y., Miwa, M., Cohen, K. B., Hunter, L. E., Ananiadou, S., & Tsujii, J. (2011). "U-Compare: A modular NLP workflow construction and evaluation system". *IBM Journal of Research and Development*, 55(3), 11:11-11:10.

Thompson, P., Kano, Y., McNaught, J., Pettifer, S., Attwood, T. K., Keane, J. and Ananiadou, S.. (2011). Promoting Interoperability of Resources in META-SHARE. In *Proceedings of the IJCNLP Workshop on Language Resources, Technology and Services in the Sharing Paradigm (LRTS)*, pp. 50-58