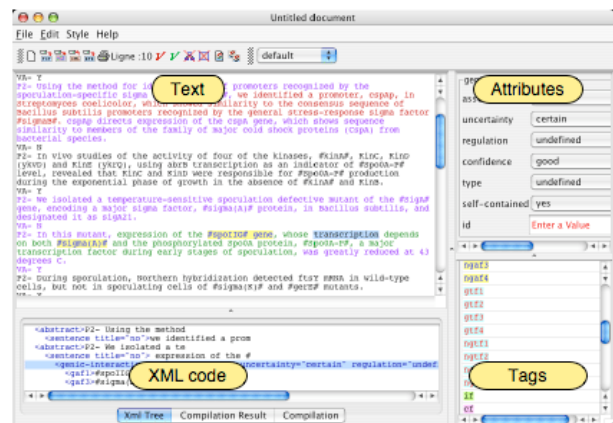## CADIXE : an XML annotation tool

Currently, there exists a lot of implementation of XML editors but most of them are oriented toward the *document creation or modification* and are poorly adapted to the *annotation task*. In this process the user wants to add progressively some XML tags to express the semantic of an existing document according to a given DTD. The CADIXE editor has been developped to fullfil this needs and to allow the user :

- To add the tags without having any constraint on the insertion order
- To visualize the document not as a tree or an XML document, but just as a classical text
- To set up the editor to his/her specific need
- To avoid a long and complex learning curve

The first requirement is of course crucial since during the annotation task the user (in our project the biologist) interprets « in real time » the meaning of the document. So it is not possible to force him/her to enter the tags in the way they are organized in the DTD, because the first element identified by the user is not always (never in fact) the root of the DTD. So, an annotation editor must provide some freedom to the annotator in the way he/she inserts the information.

The interface of CADIXE editor we developed in the Caderige project (http://caderige.imag.fr) is composed of four main parts. The *text* zone where the user sees the text he/she want to annotate, the *attributes* zone allowing to enter or to modify the values associated to the current tag (i.e. the tag in which the text cursor has been moved), the *XML code* currently generated and the list of the tags that can are allowed at a given time. Thus, the tagging process is simple : first the user highlights with the cursor the part of the text to annotate. The *tags* list shows all the relevant tags. The user

selects the tag he/she wants to apply and then enters the values in the attributes editor. There is also a mode in which the editor forced the user, through a modal dialog box, to enter the values of the attributes. This mode is relevant when the user is not familiar with the DTD.

To each tag of the DTD, the user can associate a « style sheet » allowing to express the way this tag, once inserted, will be displayed in the text. For instance, in the biological domain an annotated sentence could look like this one, each color corresponding to a specific information :

A low level of GerE activated transcription of CotD by GerE RNA polymerase but in vitro

The CADIXE editor has been implemented in JAVA 1.42 and it can be used on the following platforms : Windows, Linux, MacOs X. It is used in our project but also at the SIB (Swiss Institute of Bioinformatics) with another DTD, in the frame of the European project « BioMint ». The software can be (freely) downloaded by any researcher which is involved in a document annotation task, if you are interested just contact : gilles.bisson@imag.fr.