

Why MMAX2 Is Not Free

© Christoph Müller
EML Research gGmbH
<http://mmax.eml-research.de>

25th August 2004

Recently, EML Research gGmbH made public that a license fee will be charged for the new version of the annotation tool MMAX, i.e. MMAX2. While some people signalled understanding for this step, others reacted differently. The main objection against *selling licenses* (as opposed to *giving the software away*) was that basic research relied on free software resp. that only free software fosters basic research. In the ensuing discussions, we found ourselves telling the basically same story over and over again. So we decided to put together a short document explaining why MMAX2 is not free. Please feel free to send your comments or questions to mmax@eml-research.de. We would like to hear from you.

Background The current version's *predecessor*, MMAX, was developed in the context of the BMBF-funded research project EMBASSI (<http://www.embassi.de>). This tool was made (and still is) freely available, and its success (in terms of research projects using it)¹ was considerable. Although this tool was pretty limited and mainly tailored for coreference annotation, some of its features (in particular, its customizability) put it above most other tools available at the time. The main shortcomings were poor display speed and visualization, limited types of markable relations, and of course the limitation to only a single level of annotation. The basic concepts behind MMAX, however, appeared to be right. So, we decided to do a makeover, which soon turned out to be an almost complete reimplementation. The idea was to build on and extend the advantages of MMAX, while at the same time removing its limitations and weaknesses as far as possible. The result of this was MMAX2.

Current State Until now, the development of MMAX2 was covered by our own budget. In order to continue financing, we tried to participate in several research proposals for annotation-related projects as minor project partners, but this was not (yet) successful. The tool itself is now in a beta phase, and it is stable enough to be practically usable for 'real-world' annotation. It does also implement all necessary functionalities. But development is not yet complete: there will of course be bugs to be fixed, lesser functions to be added, some things will have to be optimized, and so on. More

¹Cf. the list of reference projects at <http://mmax.eml-research.de>

importantly, however, new *features and functionalities* will have to be added as well: as a major addition, we think of an API which will allow users to add their own plug-ins to the tool. This will enable users to modify and extend MMAX2 to their own needs. We think that this approach (i.e. creating and maintaining a well-defined interface and supporting people using it) is a good way to allow for the tool's customizability. It will also be through this interface that components for (semi)automatic annotation can be integrated.

This API is only one of several 'visions' we have about what MMAX2 could be like in the near future, and we would like to put all of them into action. At the same time, we think that there is good reason why we should do this ourselves.

The first reason is a merely technical one: As the original developers of MMAX and MMAX2, we know its inner workings and have a pretty good idea how to implement things in a reasonable way. This is particularly important for things like the API, which accesses the very core of the tool, but it holds for other features as well. This is not to say, of course, that other people, given sufficient time and support, could not also find their ways through the program. But in our opinion this type of work is undesirable and very inefficient and thus expensive.

The second reason is that we have actually gotten to like working on the tool, and judging from the feedback we got, what we have done so far does not appear to have been too bad. Therefore, we would like to continue this work. It is also for this reason that we do not intend to make the tool open source.

FAQs

I could write my own program. Why should I use MMAX2? In the past it was common practice (and it still is today) to write a specialized annotation tool for any given annotation task from scratch. This is often motivated by the simple lack of an appropriate and sufficiently customizable tool. Due to their high degree of specialization, tools produced this way are often of very limited scope. They often do not scale nicely if e.g. half way through an annotation pilot study, some scheme modifications become necessary. At the end of a project, they are normally discarded, and new tools are written for the follow-up projects. Annotators who may have been familiar with the previous tool have to be trained on these new tools again. MMAX2, in contrast to that, was not developed in the context of a particular research project with a limited scope. This made it possible to see the task of annotation in a broader sense and to strive for something like generality. Supporting MMAX2 (in whatever form, cf. below) means being able to concentrate on the actual research. And also in plain financial terms, using MMAX2 is much cheaper than developing your own tool.

What is the preferred way of obtaining a license? If you are interested in using MMAX2, there are several ways to do so. The preferred way is to make EML Research gGmbH a minor partner in your project. As a project partner, we will be able to not only supply the annotation tool (including support), but also to help in customization and other tool- or data-related questions. Of course, this scenario is only applicable if a non-trivial amount of customization is required. If, on the other hand,

the off-the-shelf functionalities of MMAX2 are sufficient for your annotation tasks, simply buying a license is preferable. If you are still in the process of writing your project proposal, costs for purchasing specialized software can be included there. The license model is such that a site license with an unlimited number of users should be the best deal here. If, on the other hand, your project has already started and not much money has been allocated for such expenses, the license model does also provide for restricted but much cheaper licenses.

Where does the money go? The money that we ask for MMAX2 licenses is not intended to compensate the cost of its development. Neither does this money constitute any profit for any of its developers or EML Research. Instead, its sole purpose is to finance continuous maintenance, support, and in particular further development of MMAX2. At EML Research gGmbH, just like at any other not-for-profit research institution, we are supposed to acquire third-party funding to complement our budget and to justify our work. Thus, technically, money paid for MMAX2 licenses will be treated as third-party money. This is also important because one factor in the NLP group's performance evaluation is the amount of third-party money acquired. Also in this respect, EML Research gGmbH is not different from other institutions.

Will I be dependent on MMAX2 if I use it to create annotated data? One of the arguments brought forth against selling user licenses is that in order for some kind of data format *standard* to develop, software supporting this format should be free. In contrast to this, we think that it is important that the data format itself is 'open' in the sense that it can be inspected and processed with other programs as well. This is true for both the data format used by MMAX and MMAX2. Data is stored in a documented representation format implemented in standard XML. This data can be used outside the tool itself, either by means of the Discourse API or some standard XML processing suite. Also, it is planned to release a *free MMAX2 Viewer* for browsing and querying annotated MMAX2 data.