Reviewer's report

Title: e-MIR2: a public online inventory of medical informatics resources

Version: 1 Date: 21 March 2012

Reviewer: Thomas Deserno

Reviewer's report:

eMIR2: A public online inventory of medical informatics resources

The authors describe the structure and creation of a web-based database that hosts online resources in medical informatics. Records of databases such as Pubmed are processed automatically, and links and meta information are extracted according to a classification scheme defined for this purpose. The web site is alive, and currently 282 data entries are available.

Although this reviewer agrees in the need of such a resource, he disagrees in some of the methods chosen by the authors. The paper needs major revision before it can be accepted for publication.

1. Improve taxonomy: Figure 1 shows the taxonomy developed for the eMIR^2 project. The functionality axes mix nouns and verbs, with is inconsistent to the description provided in the text. For instance: education -> educate, etc. in particular, this holds for the second level of the structure.

a. The paper should briefly define the terms. What is meant particularly when differing “Software/application” and “Software/Standalone”? The terms within the levels of hierarchy should be from the same level of semantics, and ambiguities and overlap should be avoided.

b. With respect to the methods, the terms and classification scheme should be developed manually, and then automatically terms extracted from title and abstract can be mapped.

c. When defining the terms, state of the art and other terminologies must be taken into account. Here, the authors can learn.

d. On page 4 you state “we searched existing taxonomies…”, specify and name them!

e. On page 4 you state “all of the existing resource taxonomies…” again, specify and name then including references. What about [1]?

2. Process more data: Since the database is filled automatically, much more papers should be processed.

a. I disagree in the selection of just some of the MI journals. The authors shall not define just another subset, but refer to existing ones. For instance, see [2] for MI journal sets.

b. Why do you select journals at all and do not process all abstracts?
c. It seems that within the last months, nothing has been added, although the paper states that the database is updated regularly.

d. Give more details on the training set (page 5). Why 204 papers, which papers and how is this set composed? Why is it representative?

e. Important resources cannot be found. The entire field of medical image processing, including open resources in terms of numerous databases, software packages, libraries, and case collections is missed. Note that this indicate the failure of the concept.

3. Provide evaluation: The database results form automatic processing of bibliometric entries. You should present an evaluation.

a. For instance, on a defined set of papers count the number of false positives and false negatives. From 9492 to 282; how many relevant papers were missed and how many have been excluded manually, before ending up with 282?

b. Define important resources, e.g. from NIH, CaBIG, ImageJ, etc, and evaluate that these are caught with your processing.

c. Some of the 282 papers – at least in my opinion – are on different content than describing online MI resources. For instance, “Biomedical Informatics Publication” – what type of resource is meant here?

4. Improve the interface: The web interface is inconvenient in use, and required scrolling since the usual wide screen monitor resolution is not acknowledged.

a. How to link one resource with several papers?

b. How to link multiple resources described in one paper, e.g., in a review?

c. How to handle versions of the same resource?

d. How to use at least AND and OR when searching your database? The user interface currently is very poor.

5. Discuss your work self-critically: I'm wondering, for instance, why only 20 links form 16 papers have been extracted. Since you aim at finding online resources, this indicated again that the method is not working at all.

References:


Level of interest: An article of importance in its field

Quality of written English: Needs some language corrections before being published
**Statistical review:** No, the manuscript does not need to be seen by a statistician.