Reviewer's report

Title: Classifying the Precancers: A Meta Data Approach

Authors:

Dr Jules J Berman (bermanj@mail.nih.gov)
Donald E Henson (hensond@comcast.net)

Version: 1 Date: 13 May 2003

Reviewer: Catherine A Smith

Level of interest: A paper whose findings are important to those with closely related research interests

Advice on publication: Unable to decide on acceptance or rejection until the authors have responded to the compulsory revisions

COMMENTS

*Compulsory Revisions*

1. This appears to be a comprehensive writeup of a presentation made by the first author at the APIII conference in October 2002. Since this presentation was "published" as an abstract only and not in its fullest form, this is the sort of "duplicate" publication that is perfectly acceptable, but it would be nice to alert the reader to this so that they realize the two are one and the same.

2. page 5, paragraph 5:

"clasification" instead of "classification".

3. page 7:

Point #3 head needs to conclude in a colon like its fellow points.

4. page 8:

I think there is some glossing over of the difference between the Unified Medical Language System itself and the Metathesaurus (MTH) which is one of the features of the system. It is unfortunately common to treat the two as if they are one and the same, which can serve to cloud the Metathesaurus' function as a thesaurus that serves to unify thesauri. The authors have an opportunity to avoid this problem.

5. page 8:

The authors are proposing a badly needed classification for which they have offered ample and convincing justification. In doing so, they have proposed a new terminology. I believe that it is customary in proposing a new terminology to give some idea of the raw material from which it was developed. I think it is even more important in cases where the new terminology is being compared against concepts contained in the UMLS Metathesaurus (MTH), which itself is an overarching collection of terms from multiple vocabularies and classifications. Otherwise, it is difficult to tell how
much of the new terminology is new, and how much is a re-sorting of existing terms.

The authors indicate in their Discussion section that they recognize this as a potential limitation, when they cite the "issue of omission" inherent in this or any project being the work of two people. It is simply difficult to judge how representative the new terminology is which is being proposed when the source of the terms is unclear.

To further clarify this, I would recommend adding some explanatory material as to the source of the terms used for the precancers.

6. page 8:
I would recommend clarifying whether these new terms for precancers were found to be matches to existing MTH concepts, to existing synonyms for existing MTH concepts, or both. This again addresses the problem of how representative the terminology is and of what concepts.

*Discretionary revisions*

1. page 3, paragraph 3:
The use of the term "meta data" with a space between meta and data is unusual. It's the form of the term used in the original W3 standard but certainly not the common parlance.

2. page 3, paragraph 3:
The authors have identified 4 "flanking tags" which are recognizable to me as drawn from the UMLS Metathesaurus. I am not sure that they will be instantly recognizable to every reader of this paper; "cui" for "CUI" (Concept Unique Identifier) in particular makes some demands on previous familiarity with the System. I would suggest clarifying the meaning of these flanking tags at this point in the paper, rather than later on, since their meaning is important to make clear for the rest of the paper to make sense.

**Competing interests:**

None declared.