

Discussion of Kees Woestenenk's comments on my PhD thesis

Reinout van Rees

Mostly written on 7 January 2007, packaged up on 16 January 2008

1 Introduction

After publication of my PhD thesis [1], Kees Woestenenk published a set of comments [2] *because this thesis has so many inaccuracies that it might hinder other developments that are in progress. . . . information that is available for academics and researchers should be reliable, which is not the case if they rely on the information in this thesis.* Kees hopes that his comments can reduce the damage I'm doing with my thesis.

A thesis being discussed is way better than a thesis that is slowly gathering dust on some shelves. The academic world thrives on discussion and on the sharpening of ideas against each other, so I was glad that Kees put his opinion on paper and circulated it to some members of my committee.

My initial summary of Kees' comments: my research of existing developments (and specifically the STABU LexiCon) has many inaccuracies and plain errors; my conclusions about those developments are way too critical and don't take into account the amount of work that has been invested in those developments and standards; the new possibilities of the internet don't offer anything new over existing well-proven technologies like ISO-STEP.

2 Structure of this document

First, especially in his discussion of chapter 3, Kees points out many inaccuracies and factual errors. There is no real alternative apart from listing the allegations and addressing them. However tedious reading the resulting section will probably turn out to be.

Second, the scope of my PhD thesis and the scope of a PhD student's work seems to be misunderstood.

Third, I will discuss the differences in the, if you will, 'business plan' apparently preferred by Kees and the one expressed in my thesis. I would like to see my ideas bear fruit and Kees wants to see the LexiCon become succesful: is there a difference in the chance of market acceptance?

3 Alleged inaccuracies and errors in chapter 3

3.1 Specification and classification

In the discussion Specification is restricted to the document used for tendering between client and contractor, and more specific to tendering, subcontracting and purchase orders are not discussed.

At the bottom of page 27 I say: **One common way of subdividing the textual Specification is subdivision into parts called chapters. Traditionally chapters often correspond with branches of the industry or types of work.** Subcontracting is often done on the level of industry branches. All the paintwork is subcontracted. It is implied that you can take the one chapter on paintwork and give it to the painting subcontractor.

The desired match between specification and costs estimation (p 28) is just the reason why the information in specifications is ordered as it is.

I explicitly say at the top of page 28 that a subdivision in crafts **makes it easier to provide a cost estimation by allowing different experts to estimate their part.**

What is really meant with actual Specification units (p 28) remains unclear.

See figure 3.3 on page 33: I consider both ‘SpecItem’ and ‘SpecItem-Part’ to be the actual specification units (or items) that contain the actual information on doors, paint, etcetera. The point is that you can use the classification hierarchy of the specification to order information, but you always end up with the subsection ‘doors’ instead of actual door items.

The ETIM system (p 30+) is described only as a classification system. The author does not mention the specification templates which form the real content of the system and which are used in the process of searching and purchasing electrical products.

On page 31 I say that ETIM sees **very widespread use in the electrotechnical and installation sector, mostly as a basis for catalogs and catalog lookups.** So searching and purchasing of electrical products.

ETIM does not describe installations, only products.

Which is exactly as I say (page 32): **A second restriction is that the Classification only supports single components. As long as it is one item that you can buy, the item fits in: an electrical socket or a light switch. Complex items that consist of sub-components (like a heating system) do not fit in.**

The description of the STABU Specification system is incomplete in that it does not mention the construction parts which are the subject of the specification.

Again, see figure 3.3 on page 33. I’m pretty sure that ‘construction part’ is what I call ‘SpecItemPart’ in that *simplified* diagram. I worked from

Dutch system design documents, so perhaps there are other, more common, translation options.

The author does not mention the possibility to group the specification items according to the user's wish (the Nl-SfB system included) is not mentioned either.

Indeed, I did not mention this. From what I gathered (late December 2006) from my visit to STABU when delivering a box of theses, they are currently very busy to use/enable this possibility. So at the time of writing, the Nl-SfB work had not been done by STABU itself yet. I doubt (I don't have figures) that many companies add their own classification, most will just use the build-in one.

Being database oriented, and being composed of specification units that can be combined in several ways, the STABU Specification system is not so much paper oriented as is suggested (p 34), but could be extended to an object oriented information system, by using semantic attributes instead of the current text attributes.

It is precisely because it are purely textual attributes that I call it a paper oriented system. The theoretical possibility to re-work the system into a data-oriented system has no influence on the *current* paper oriented nature. Reworking it into a data-oriented system is a good idea. **A system that exposes the real content, instead of a representation, will prove more useful and valuable** (page 34). In fact, such a reworking of a large number of the BC applications is what I propose. Make the data available. A difference in approach is probably that I propose to make the data available with just the semantics of the individual applications and then to start trying to wire the applications together. If my conjecture is correct, the LexiCon approach first wants to have a world-wide accepted LexiCon that provides all the semantic needs of BC before the applications should be changed.

3.2 PDT

The treatment of ISO-STEP is superficial and does not give enough credit to the enormous effort that is invested in the development of this set of standards.

Neither do I give explicit credit for the enormous effort that went into many Internet technologies, for instance. That people invested a lot of effort into something does not by definition make it better than other technologies.

The comment of 'giving enough credit' seems to be tied to the comment at the end of Kees' document: *It just does not make sense to set these developments aside and re-invent the wheel.* I apparently should 'just accept' the current research efforts out of a sense of proper respect to the amount of effort already expended. I must take exception to this: a mindset like this has nothing to do with critical scientific research (which is what a PhD is

supposed to do).

The description of ISO 12006-3 (section 3.3.3) completely overlooks the concepts of Relationships and Collections, which are vital parts of this model. The statement that this model has only two structuring mechanisms: inheritance and composition (p 38) is simply incorrect, see the types of Relationships and Collections.

Relationship is an abstract base class for inheritance, grouping and several others. As such, it is not a core *structuring* mechanism for organising the *objects*. When structuring doors and floor tiles, the core way of structuring them has always been the inheritance relationship, followed at a respectful distance by the composition relationship. At least, that has been the practice for the LexiCon.

The author observes some weaknesses in ISO 12206-3. The first one, the possibility to refer to external information sources does not make sense, either a reference is made to unstructured information, or a mapping can be made between structured information, both are possible. The problem of interconnecting libraries based on ISO 12006-3 is also a mapping problem that can be solved.

My point is that this mapping problem cannot be solved when you adhere purely to the ISO specification. The model abounds with internal relations, but only a text field can be used to refer something that's external. So compared to the rigor of the internal references, semantically rich external references are (virtually) non-existent. That's a closed-world assumption.

A mapping problem that can be solved? That cannot be done with a textual reference to an item in another 12006-3 database. The standard will need to be changed to allow it.

The STABU LexiCon ... with a built-in multi-inheritance taxonomy ...

The use of multi-inheritance is a very recent addition (as of late 2006, when the thesis was written) to the tool. The last known version of the LexiCon, at <http://www.stabu-lexicon.nl/>, was converted to html by myself: there was no multi-inheritance. For a long time, multi-inheritance wasn't even possible in the LexiCon editing tool. And not allowed.

The statement that only the inheritance mechanism is used (p 42) is not true: the LexiCon uses all types of Relationships defined in ISO 12006-3.

The statement is true, sadly. I qualify it on that page 42 with 'the most recent version', which was the one at <http://www.stabu-lexicon.nl/> at the time of writing. The all-important object composition relation isn't used, but reserved for future commercial extensions. Properties are attached, but they aren't what I'd call real object properties. So not 'length' and 'fire resistance' but abstract 'structural = true/false'-like properties that only serve to create the top-level subdivision in the first abstract levels of the inheritance hierarchy.

There was no "lack of data format" (p 43) ...

I explain that 'data format' means a format for the instance data, not for

the taxonomy data. The LexiCon contents itself can be exported to XML just fine, but there is no format in which you can use the LexiCon to talk about your catalog or your specification. (Which is what eConstruct tried to address with its bcXML format, as mentioned on that page 43).

Contrary to the author's statement, the lexicon does offer the possibility to attach properties.

Yes. Correct. I do think I mention it correctly elsewhere, but the way I say it here is an error. (I corrected it in the PDF/html version on my website).

3.3 Personal comment on this section

I worked on this document on 7 and 13 January 2007 (just days before my thesis defense). I finished it up to this point, but severely disliked doing something negative like this. So I didn't continue with it, hoping to find a different form for it.

On 16 January 2008, one year and a day after my succesful defense, I decided that I'd just finish it off and be done with it. Kees pointed out errors in my document; I point out errors in his document.

I've added some short comments to the two, already planned, sections below to end on a more positive note.

4 Scope of the thesis and of a PhD's work

A PhD student should do innovative research. A PhD student should not make small incremental improvements to existing technologies, as seems to be the desire of Kees Woestenenk. The LexiCon technology, apparently, is the perfect solution once a few remaining small issues are worked out.

I have done background research, as is expected of a PhD, into the fundamentals of information exchange (in the BC industry). And in doing so, I discovered problems that hamper the LexiCon and more specifically the LexiCon's chances of completion and adoption.

In such a case, small incremental improvements to the existing technology is not what is expected of a PhD student.

5 'Business plan' and technology success

5.1 Boil the ocean

My impression is that the LexiCon is basically a 'boil the ocean' [3] approach: something that is way too ambitious, effectively impossible. Everything has to be done in one go: the content of the LexiCon has to be complete, all applications must support it, all users must accept it. The problem:

- The content must be complete, though this is impossible to achieve without funding by the users that the system can't have before it is complete. Chicken/egg problem.
- All applications must support it, though the application providers don't have users yet to fund it and though they must trust this to be the ultimate system once they've funded it to completeness. See [4].
- All users must accept it before there are applications to demonstrate the feasibility. And they must fund the development 'as it is for the good of the industry'.

The approach that I advocate attempts to solve those problems in a number of ways.

- Development in small portions. A limited ontology started in one sector of the industry can start to attract users beyond the initial working prototype. Relatively simple to fund with a good possibility to recoup the costs. The possibility of integration into a greater whole remains wide open.
- Even an initial prototype application ought to provide business value. The more usage, the more business value. This means a positive feedback loop which is almost the anti-thesis of the 'boil the ocean' approach.
- An initial application (for a few companies, limited to one sector) will be developed together with the users (that help to build their own limited ontology with their domain knowledge). Funding can take place out of well-understood private interest: business value.

'My' approach has practical problems of its own: I'm sure they're solvable with a decent amount of practical organising and application of brain power. I'm equally sure that the current (well, 2000-2005) LexiCon approach has unsolvable basic problems. Sorry for that.

References

- [1] Reinout van Rees. *New instruments for dynamic Building-Construction: computer as partner in construction*. PhD thesis, Delft University of Technology, January 2007. Available on-line at <http://vanrees.org/research/phd>.
- [2] Kees Woestenenk. Thesis of Reinout van Rees, 2006. Comments by Kees Woestenenk, 2007-01-02, January 2007. Available on-line at http://www.ifd-library.com/data/Thesis_vanRees_Comments_2007-0102.pdf.

- [3] Boil the ocean, June 2004. Available on-line at <http://www.bobcongdon.net/blog/2004/06/boil-ocean.html>.
- [4] Jasper Feenstra. LexiCon expert panel discussion, October 2004. Available on-line at <http://vanrees.org/weblog/1098802937>.