

COMP700 Resources: Dealing with References

version 1.0

Dr. C. Maria Keet

email: keet@ukzn.ac.za, home: <http://www.meteck.org>
School of Computer Science, University of KwaZulu-Natal

April 18, 2011

Contents

1	Introduction	1
1.1	Plagiarism and borderlines	2
1.2	Quoting versus citing	2
1.3	The remainder of this document	3
2	Referencing in text	3
2.1	Referencing modes	3
2.2	What to put where	4
3	The “References” section	5
4	Managing your references	6
5	What you can (not) cite	8

1 Introduction

When working on a project or a thesis, one may, at times, feel as if one is working alone in isolation. This may be true for various reasons, but ought not to be the case, unless you are working on a PhD thesis when you have to demonstrate you have what it takes to go at least one whole turn through the knowledge creation spiral¹. What certainly holds, is that

- the *topic* of your project or thesis does not exist in isolation, and
- you are not set out to work so as to reinvent the wheel.

Put differently: *you are building upon other people’s contributions*. That is how science and engineering move forward.

Building upon other people’s results entails that you are aware what they did and understand the pros and cons of those earlier proposals. This *must* be reflected in the article or thesis you are or will be writing. In this document, we focus on how to go about giving credit where credit is due; not doing so amounts to **plagiarism**, which is the sin of sins in academia. If you plagiarise during your thesis writing, you will be reprimanded officially, if you have done so in your thesis and it is found out after graduation, your degree will be revoked, if you do it as a scientist, you will be fired.

¹If the latter is the case, then you should go back to work now, instead of reading here what you should know already.

1.1 Plagiarism and borderlines

Plagiarism is to copy whole sentence(s) from another document without giving a reference to the source. ‘Copy’ here is taken liberally. So, if you were to take the original sentence from [4]

On average, those who commenced with a foundational ontology added more new classes and class axioms, and significantly less object properties than those who started from scratch.

and copy it without putting it in indentation as quote or between quotation marks (“”), it is plagiarism. Or when you make only very minor changes, e.g.,

On average, those who started with a foundational ontology added more new classes and class axioms, and significantly less object properties than those who started with a *tabula rasa*.

without adding a reference to [4], then it is still plagiarism, even though the two strings are not exactly the same. True, this is difficult for mathematically-oriented papers when one has to introduce the syntax and semantics of a language that was already presented in another paper, so there are exceptions for those paragraphs (but try to fiddle a bit, where possible), provided the reference to the paper that introduced the language is given (e.g., [2] reuses the *DLR_{US}* language introduced in [1], duly referenced).

1.2 Quoting versus citing

Longer quotes, like the one from [4], above, should be *indented*, whereas shorter pieces can be put in-line in quotation marks, e.g.:

We are only interested in the “significantly less object properties” of [4] instead of the whole gamut.

You also can emphasise something in a quotation, but this has to be indicated as such. Here is an example of a longer, indented quotation and emphasis:

Against expectations, Keet [4] concluded the following from her experiments:

On average, those who commenced with a foundational ontology added more new classes and class axioms, and *significantly less object properties* than those who started from scratch. (emphasis added)

Now, let us take a look at that peculiar aspect of it.

Quotations have to be truthful to the original. For instance, if the original sentence were:

Ontologists tend to be outspoken about the usefulness of foundational (top-level) ontologies, such as BFO, DOLCE [1], GFO [2], and SUMO: either they are perceived to be essential or an impractical burden.

and you quote it as:

Despite our own experiences in developing domain ontologies, Keet [4] dares to claim that “... foundational (top-level) ontologies ... are ... essential.”

then this is clearly *not* truthful to the original sentence. It is permitted to skip *irrelevant* parts of a larger piece of text as long as they do not modify the overall message of the original.

Fiddling a bit with the layout to correct style (captialization) or substituting a relative pronoun with the original noun is permitted, which is always put between square brackets; for instance:

The dispute was succinctly summarised as “either [foundational ontologies] are perceived to be essential or an impractical burden” [4].

Instead of a quotation, which is fairly common in the humanities but not in computer science (because it takes up a lot of space), one can paraphrase it. For instance, you can put in the text

Keet [4] demonstrated that the 52 test subjects added, on average, more new OWL classes and class axioms, and significantly less object properties when they started with the OWLized DOLCE.

or

It has been shown that more new classes and class axioms were added when ontology developers started with a foundational ontology compared to those who did not [4].

In these two cases you *cite* material, but do not *quote* it.

1.3 The remainder of this document

In the remainder of this document there are a few guidelines on how you can refer to the references, the references section itself, managing references, and what is (not) referencable.

To reiterate: a section with references is *essential* to the report/article/thesis.

2 Referencing in text

You have seen in the examples in the previous section the numbers between square brackets, like a “[1]”. This is just one way of putting references in the main text—the simplest one that is computationally manageable, in fact.

First, there are two principle *ways of referencing* in the text: footnote/endnote or names/numbers. Second, within each option, there are minor style variations. Either way you choose: **be consistent** throughout the text.

2.1 Referencing modes

The first way of referencing, footnote or endnote, is very common in the humanities when it comes to ‘real’ references. For computer science, the may be common for Web references *only*, and even then it depends on the outlet which one is preferred (and due to page limitations, one may take up less space than the other, hence make the difference). An example of footnote referencing is the following.

What constitutes a nonviolent personality? Pontara describes 10 characteristics that a person with a nonviolent personality should possess to a great extent². One of these is mildness, where he asserts in §3.2.7 that “[a] person equipped with a nonviolent personality is not outside the political struggle”³, which I will discuss in detail in the remainder of this essay.

²Pontara, G. (2011). The nonviolent personality. In: Keet, C.M. (Ed.), translated from Italian *La personalità nonviolenta*. 54p. Online: <http://www.meteck.org/files/NonviolentPersonality.pdf>. Last accessed: April 15, 2011.

³*op. cit.* p27.

That is, the first reference to the document by Pontara [6] is referenced in full, the subsequent one refers to the previous note and contains only the page number. The full citation is normally also included in a separate “References” section at the end of the text.

The ‘cs referencing with footnotes’ is not always friendly to the originators, because references count in citation indices (the more your paper is cited, the higher the impact of your contribution to the field), but footnotes/endnotes do not. For instance, it is possible (but again, not nice to your colleagues) to do the following

We developed our ontology in Protégé 4.1beta⁴ by first importing the taxonomy of part-whole relations⁵.

Very nice would be, instead:

We developed our ontology in Protégé 4.1beta [3] by first importing the taxonomy of part-whole relations [5].

Citing software is, to a large extent, a judgement call. Protégé is so widely known within the ontology development community so that, if you write for *that* community, one would not cite it at all anymore. Surely one would not cite—paper or URL—say, Linux OS in your materials & methods section (unless there’s a very recent flashing new earth-shattering feature). New, or for the target audience unfamiliar, software applications should be cited so that a reader can follow up on what you have done.

You have already come across the other referencing mode, because it has been used throughout this document. More specifically, I used the numbering scheme. There is just one section of references, which are numbered alphabetically in this case, and that’s it. The previous example with a naming scheme could look like this:

We developed our ontology in Protégé 4.1beta [GMF⁺03] by first importing the taxonomy of part-whole relations [KA08].

or

We developed our ontology in Protégé 4.1beta (Gennari *et. al.*, 2003) by first importing the taxonomy of part-whole relations (Keet & Artale, 2008).

But the final layout depends on which package you use, and which settings, which will receive attention in Section 4.

2.2 What to put where

The examples in the previous section already introduced several options for how to include a reference in the text. Here I shall list that more systematically.

What has most effect on the formatting is if you choose a numbering or a naming scheme. For instance, this is good:

Keet and Artale (2008) developed a taxonomy of part-whole relations.

but if the reference were a number, rendering the sentence like this:

[5] developed a taxonomy of part-whole relations.

then that is not acceptable. In the second case, you would have to write

⁴<http://stanford.protege.edu>

⁵<http://www.meteck.org/files/ontologies/pwrelations.owl>

Keet and Artale [5] developed a taxonomy of part-whole relations.

or, if you refer to more details, then the reference can be put at the end of the sentence:

Keet and Artale developed a taxonomy of part-whole relations, which has as major division ‘real’ parthood relations versus those relations that only motivated by linguistics might seem a parthood relation, but are not [5].

When using the naming scheme and there are more than three authors for a single reference, you do not include the whole list of authors in the text, but abbreviate it with *et. al.*—in italics, and with the dots. We have seen this in an earlier example already, e.g., with the reference to Protégé:

... ontology in Protégé 4.1beta (Gennari *et. al.*, 2003) by first...

You also can combine references into one list, where appropriate. For instance,

... with some recent results on part-whole relations that also build upon temporal logics in general [1, 2, 5].

Observe two things in the above example: they are all within the same square brackets, and in numerical order. The latter is not required, but neater.

3 The “References” section

The list of references goes at the end of the text. There is no consistency to do that always after the main text but before the appendix, or also after the appendix. Check other theses/papers for that, or, if you have an official template, adhere to the template. Regardless where it ends up after the text, the references in the “References” section have to be **consistent** and **complete**.

Whichever referencing scheme you use, *do the layout the same for the same type of entries*: if you do book titles in italics, then do so for all of them, if you include the total amount of pages, do so for all of them, and so forth.

They also have to be complete to the extent that a reader should be able to find that paper. For instance, [3, 5] are complete entries. Conference and journal abbreviations and cutting down a long list of authors to one or three authors with an “*et. al.*” may be acceptable, too (and in some outlets even required), e.g.:

GENNARI, J. H. *et. al.*. The evolution of Protégé: an environment for knowledge-based systems development. *Int. J. of Hum.-Comp. Studies* 58, 1 (2003), 89-123.

may be fine, but *not* something like

GENNARI, J. H. *et. al.*. The evolution of Protégé: an environment for knowledge-based systems development. 2003.

because it misses the publication venue details, nor

GENNARI, J. H., MUSEN, M. A., FERGERSON, R. W., GROSSO, W. E., CRUBÉZY, M., ERIKSSON, H.. *Int. J. of Hum.-Comp. Studies*, 1 (2003).

because it misses the title of the paper, volume of the journal and page numbers, and has only several authors instead of all (or: misses the “*et. al.*” after the last-mentioned author).

References to conference papers have their own acceptable or tolerable ‘shorthand notations’. For instance, if the authors were really short on space, then a complete reference like [1] can be seen to be abbreviated as:

ARTALE, A., FRANCONI, E., WOLTER, F., AND ZAKHARYASCHEV, M. A temporal description logic for reasoning about conceptual schemas and queries. *Proc. of JELIA'02*, Springer LNAI 2424, 98-110.

But again, this is only by exception and is done only when the readership is familiar with those venues. You, as a budding scientist reading those papers, may not be—and that is the only reason why the two examples of ‘condensed’ references are described here.

Nevertheless, this raises the question: What are the ‘minimal’ components of a reference? This depends on the reference *style* you use. For instance, there is a so-called “Harvard referencing” and “Chicago style” that have elaborate guides with the dos and don’ts, there are common styles in computer science, such as of the IEEE, Springer LNCS, and ACM, and journals tend to have their own requirements how to reference each type of resource. You’d be crazy to read through all those guidelines and adapt your references each time you have to follow another scheme. Scientists and software developers got together, and developed reference management software to do this for you, so, unlike other resources on referencing (and there are a lot of them), I will not waste time re-writing a guide on that, because we can get that sorted out automatically. This is the topic of the next section.

4 Managing your references

When you conduct your literature research to read up on the topic you have chosen for your thesis, article, or technical report, you will come across many resources, some more useful than others. Of those deemed relevant, you should note the publication details immediately to save yourself from wasting time searching for that resource again in a few months’ time. What information about the publication venue of that resource should you record? We have seen in the previous section that the notion of ‘complete’ information might, in fact, vary slightly. In addition, you may have noticed that the references you have come across have different layouts. Although you are going to focus on writing your honours project now, you may wish to continue with a masters and/or, if the results obtained are really good and novel, you may be involved in writing a scientific paper afterwards, which may have its own requirements for formatting references. Manually reformatting the references is a painstaking task and prone to introduce inconsistencies in the layouts. The latter is considered *sloppyness* and not appreciated: if you obviously do not care about your text, why should the reader spend precious time to read it?

Fortunately, many have gone before you who wanted to have a nice software-based reference management system that also does automatically the reformatting of the references adjusted to the venue and in the right order. There are now several such tools around that come in three different flavours. First, there are ‘general science’ reference management tools for people who write texts in OpenOffice or MS Word, such as Reference Manager⁶ and Endnote⁷, which, however, are stand-alone applications that are not for free. Second, there are free tools that take a social networking approach to reference management, such as Mendeley⁸. The third one is for L^AT_EX aficionados, such as Jabref⁹, BibDesk¹⁰, and BibTool¹¹ (more general information at: <http://www.bibtex.org/>).

Basically, you store your references in a fancy database and each time you use a reference, you insert its key in the text. Once ready, the selected keys, your text editor, and your ‘selected export format’ get together and produce the right amount of references in the right format in the right order—automatically.

⁶<http://www.refman.com/>

⁷<http://www.endnote.com/>

⁸<http://www.mendeley.com/>

⁹<http://jabref.sourceforge.net/>

¹⁰<http://bibdesk.sourceforge.net/>

¹¹<http://www.gerd-neugebauer.de/software/TeX/BibTool/index.en.html>

- [Artale et al.(2002)Artale, Franconi, Wolter, and Zakharyashev] A. Artale, E. Franconi, F. Wolter, and M. Zakharyashev. A temporal description logic for reasoning about conceptual schemas and queries. In S. Flesca, S. Greco, N. Leone, and G. Ianni, editors, *Proceedings of the 8th Joint European Conference on Logics in Artificial Intelligence (JELIA-02)*, volume 2424 of *LNAI*, pages 98–110. Springer Verlag, 2002.
- [Artale et al.(2008)Artale, Guarino, and Keet] A. Artale, N. Guarino, and C. M. Keet. Formalising temporal constraints on part-whole relations. In G. Brewka and J. Lang, editors, *11th International Conference on Principles of Knowledge Representation and Reasoning (KR'08)*, pages 673–683. AAAI Press, 2008. URL http://www.meteck.org/files/AGK_KR08.pdf Sydney, Australia, September 16-19, 2008.
- [Gennari et al.(2003)Gennari, Musen, Fergerson, Grosso, Crubézy, Eriksson, Noy, and Tu] J. H. Gennari, M. A. Musen, R. W. Fergerson, W. E. Grosso, M. Crubézy, H. Eriksson, N. F. Noy, and S. W. Tu. The evolution of Protégé: an environment for knowledge-based systems development. *International Journal of Human-Computer Studies*, 58(1):89–123, 2003. ISSN 1071-5819. doi: [http://dx.doi.org/10.1016/S1071-5819\(02\)00127-1](http://dx.doi.org/10.1016/S1071-5819(02)00127-1).
- [Keet(2011)] C. M. Keet. The use of foundational ontologies in ontology development: an empirical assessment. In G. Antoniou et al., editors, *8th Extended Semantic Web Conference (ESWC'11)*, volume 6643 of *LNCS*, pages 321–335. Springer, 2011. Heraklion, Crete, Greece, 29 May-2 June, 2011.
- [Keet and Artale(2008)] C. M. Keet and A. Artale. Representing and reasoning over a taxonomy of part-whole relations. *Applied Ontology – Special issue on Ontological Foundations for Conceptual Modeling*, 3(1-2):91–110, 2008. URL http://www.meteck.org/files/A007_pw_AK07.pdf.
- [Pontara(2011)] G. Pontara. *The nonviolent personality*. page 54. March 2011. URL <http://www.meteck.org/files/NonviolentPersonality.pdf>. Translated from the Italian original *La personalità nonviolenta*. Online: <http://www.meteck.org/files/NonviolentPersonality.pdf>. Last Accessed: April 15, 2011.

Figure 1: The same references, but formatted with `abbrvnat`.

Let’s have a quick look at an example for L^AT_EX, which is the preferred editing and typesetting program of computer scientists. Your text goes in a `.tex` file, your references go in a `.bib` file, and your choice for reference format is a selected `.bst` file. At the end of your main file, say, `myproject.tex` you add two lines, one to specify the reference *style*, say, `model3-num-names.bst`, and one to refer to your bibliography, say `mybib.bib`, which then looks like this:

```
\bibliographystyle{model3-num-names}
\bibliography{mybib}
```

You then have to run `latex-bibtex-latex-latex` to get all references fully resolved. To produce the references for this document, I used the `acm.bst` format, and there are very many more options available that come both with a L^AT_EX editor distribution and online. Have a look at, e.g. this resource¹². For instance, when I use the very same bib file but with

```
\bibliographystyle{abbrvnat}
\bibliography{mybib}
```

the references look like depicted in Figure 1. Not just that, but you would not have seen numbers in square brackets in this document, but the authors’ names, like:

¹²<http://www.cs.stir.ac.uk/~kjt/software/latex/showbst.html>

...importing the taxonomy of part-whole relations [Keet and Artale(2008)].

You may not like the square brackets, or want to have the in-text format alike “(Keet & Artale, 2008)” instead of “[Keet and Artale(2008)]”. The .bst files are customizable, but before you have a go at that, you may want to check out the very powerful natbib package that already does most of this for you.

5 What you can (not) cite

Your thesis/article/tech report is a serious academic document. This means that you use, and build upon—hence, also reference—proper academic and reliable, material of good quality. As a rule of thumb:

- What you can reference: journal articles, conference papers, books, chapters in books (including edited conference proceedings), technical reports, other theses, standards. In general: they are *primary sources*.
- What you should *not* reference (under normal circumstances, in CS): blogs and blog posts, posts on forums, Joe Soap’s tips ’n tricks. These are, at best, *secondary sources*.

Web pages and company product manuals are trickier, but, in general, they should be avoided whenever possible. For instance, a ‘white paper’ by Company of Ruby And Perl on the Fabulously Implemented Algorithm to SCam Others may claim that their tool is the coolest around, but any company will say that about their applications. Let “[7]” be the reference to that online white paper, then you might be tempted to write a sentence in your thesis alike

CRAP’s FIASCO has the best performance [7].

A press release may say so, but your thesis is not a mouthpiece for the company. If, on the other hand, there is a paper *independently* from the organisation that compared it to similar tools and has shown it experimentally to hold, then that resource should be referenced.

Another tricky resource is Wikipedia, because it provides a nice introductory overview of many topics. However, Wikipedia is a secondary source and you should have a (reliable) textbook or handbook on those topics, in particular when it comes to your thesis. In that case, it is more appropriate to reference the textbook.

Regarding blogs, while they should be avoided as reference, if you find, say, a piece of *good* code that someone put on their blog and it cannot be found elsewhere, it is better to reference that than claiming you invented it when you did not.

References

- [1] ARTALE, A., FRANCONI, E., WOLTER, F., AND ZAKHARYASCHEV, M. A temporal description logic for reasoning about conceptual schemas and queries. In *Proceedings of the 8th Joint European Conference on Logics in Artificial Intelligence (JELIA-02)* (2002), S. Flesca, S. Greco, N. Leone, and G. Ianni, Eds., vol. 2424 of *LNAI*, Springer Verlag, pp. 98–110.
- [2] ARTALE, A., GUARINO, N., AND KEET, C. M. Formalising temporal constraints on part-whole relations. In *11th International Conference on Principles of Knowledge Representation and Reasoning (KR’08)* (2008), G. Brewka and J. Lang, Eds., AAAI Press, pp. 673–683. Sydney, Australia, September 16-19, 2008.

- [3] GENNARI, J. H., MUSEN, M. A., FERGERSON, R. W., GROSSO, W. E., CRUBÉZY, M., ERIKSSON, H., NOY, N. F., AND TU, S. W. The evolution of Protégé: an environment for knowledge-based systems development. *International Journal of Human-Computer Studies* 58, 1 (2003), 89–123.
- [4] KEET, C. M. The use of foundational ontologies in ontology development: an empirical assessment. In *8th Extended Semantic Web Conference (ESWC'11)* (2011), G. Antoniou et al., Eds., vol. 6643 of *LNCS*, Springer, pp. 321–335. Heraklion, Crete, Greece, 29 May–2 June, 2011.
- [5] KEET, C. M., AND ARTALE, A. Representing and reasoning over a taxonomy of part-whole relations. *Applied Ontology – Special issue on Ontological Foundations for Conceptual Modeling* 3, 1-2 (2008), 91–110.
- [6] PONTARA, G. *The nonviolent personality*. March 2011, p. 54. Translated from the Italian original *La personalità nonviolenta*. Online: <http://www.meteck.org/files/NonviolentPersonality.pdf>. Last Accessed: April 15, 2011.

Additional sources

Coxhead, Peter. A Referencing Style Guide. University of Birmingham. Last updated: 16 Nov 2009. <http://www.cs.bham.ac.uk/~pxc/refs/index.html> Last accesses: April 15, 2011.

The Open University. The IT and Computing Project—Resources. Faculty of Technology. version 2. 2003 [CD-ROM CDR0579].