Building deductive networks.

Ameet Sharma

June 27, 2017

ameet_n_sharma@hotmail.com

Abstract

This article builds on ideas in [3]. We propose enhancing journal papers with XML to facilitate the creation of deductive networks[3] (DNs).

1 Introduction

Recently, there was been research done on visualizing networks of research papers. Examples are [2] and [1]. These efforts are worthwhile. But as they are limited to analysis of citations, they can only be used for statistical analysis. We propose that by enhancing journal papers with XML, we can create true knowledge networks. These are called deductive networks in [3] (DNs). Such networks would be a tool for understanding and verifying the knowledge contained in the papers, especially for mathematical papers. They would also serve as an invaluable tool for peer review.

2 Limitation of citation based networks

With citation-based networks, the entire paper acts as a node. From the network, all one can now is that one papers refers to another. There's no

information about the nature of the connection between the two papers. For that type of information, you simply have to read the two papers. But if we supplement paper by defining entities within the paper, we can have networks that connect those entities instead. These "entities" would be chunks of knowledge.

3 Can the knowledge in papers be divided into chunks?

Mathematical knowledge naturally divides. We use definitions, theorems, corollaries, lemmas etc. We label these items, and then we are able to refer to them using these labels, as opposed to restating the content. So mathematical knowledge is modular in nature. This allows mathematical papers to be XML-enhanced using the methods described in [3]. But is all knowledge modular? Specifically, can we always take any paper and divide it into distinct chunks which we can then label and refer to? We can always find an arbitrary way to do this, but not necessarily a meaningful way. But in many cases we can find such chunks. Authors often demarcate such chunks by defining terms, or summarizing their conclusions. The process of summarization forces them to link to together multiple facts into single whole which can then be assigned a label.

4 Reference vs deduction

These chunks could be labeled using XML and then be used as references/citations in the same way references to entire papers are currently used. However, the real value of these chunks is to form DNs, where the chunk in the current paper does not merely "refer" to the chunk in the older paper, but there is a flow of reasoning, or deduction, from the old chunk to the new chunk. ie: understanding the old chunk is necessary to some degree to understand the new chunk. As we know a reference may have multiple uses. It's not always the case that the information in the paper being referenced is necessary to understand the current paper. So citation-based networks don't necessarily form chains of reasoning. The purpose of DNs is to form chains of reasoning. This makes it important for the chunk to have some kind of unified meaning/concept. ie: it's not just an arbitrary section of a paper.

5 Implementation

In [3], it was proposed to use a single file with embedded xml, to define the chunks in a paper, as well as the deductive links from other papers. However, it might be a better idea to use 2 different files: one which contains the text of the paper with embedded xml defining its chunks. Call this the **definition** file. And a second file which defines the links from the chunks of other papers. Call this the **linkage** file.

6 A fictional example

Suppose we have two journal papers. One called "Hobbit Biology in Middle Earth" and a second called "Hobbit Behavior in Middle Earth" which references "Hobbit Biology in Middle Earth". Suppose "Hobbit Biology in Middle Earth" has no references, and "Hobbit Behavior in Middle Earth" has only one reference which is "Hobbit Biology". We can construct a citation network based on the bibliography of "Hobbit Behavior".

Suppose this is the "Hobbit Biology in Middle Earth" article:

Hobbit Biology in Middle Earth
Hobbits are....
.....
Hobbit digestive system
.....
....end of Hobbit digestive system section
Hobbit brains
....
...end of Hobbit brains section
... end of article

Suppose this is the "Hobbit Behavior in Middle Earth" article:

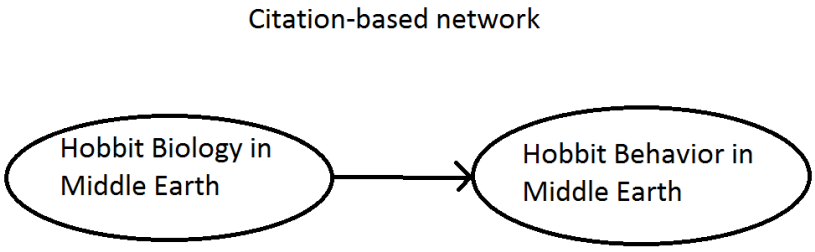
Hobbit Behavior in Middle Earth
Hobbits are....
Hobbit eating habits
.....

....end of Hobbit eating habits section
Hobbit sleeping habits
....
...end of Hobbit sleeping habits section
References
[1] Hobbit Biology in Middle Earth
... end of article

A citation network can be constructed based on analysis of the references.

(F.1)

Figure 1:



Now to construct our DN,we add XML tags around pieces of text that can be thought of as a chunk of knowledge. Let's say the definition file for "Hobbit Biology in Middle Earth" is called hobbitBiologyDefinition.knml (knml stands for knowledge markup language taken from [3]). Note that the ID is unique ID among all journal papers and knowledge chunks, designed so that any journal paper can use this ID to refer to this chunk of knowledge.

<d-knowledge>
<d-title>Hobbit Biology</d-title>
<d-id>X4329</d-id>
Hobbit Biology in Middle Earth
Hobbits are....
.....
<d-knowledge>
<d-title>Hobbit digestive system</d-title>

<d-id>X4330</d-id>
Hobbit digestive system
.....
....end of Hobbit digestive system section
</d-knowledge>
<d-knowledge>
<d-title>Hobbit brains</d-title>
<d-id>X4331</d-id>
Hobbit brains
....
...end of Hobbit brains section
</d-knowledge>
... end of article
</d-knowledge>

"Hobbit Biology in Middle Earth" has no linkage file as it does not refer to other articles. Note that we are defining the entire article as being a chunk of knowledge, with 2 nested chunks (digestive system and brains).

Similarly we construct the definition file for "Hobbit Behavior in Middle Earth" (hobbitBehaviorDefinition.knml):

<d-knowledge>
<d-title>Hobbit Behavior</d-title>
<d-id>X4529</d-id>
Hobbit Behavior in Middle Earth
Hobbits are....
<d-knowledge>
<d-title>Hobbit eating habits</d-title>
<d-id>X4530</d-id>
Hobbit eating habits
.....
....end of Hobbit eating habits section
</d-knowledge>
<d-knowledge>
<d-title>Hobbit sleeping habits</d-title>
<d-id>X4531</d-id>
Hobbit sleeping habits
....
...end of Hobbit sleeping habits section
</d-knowledge>

References
[1] Hobbit Biology in Middle Earth
... end of article
</d-knowledge>

Like the biology definition file, we have one chunk of knowledge with two chunks nested inside. Now we construct the linkage file for "Hobbit Behavior in Middle Earth" (hobbitBehaviorLinkage.knml)

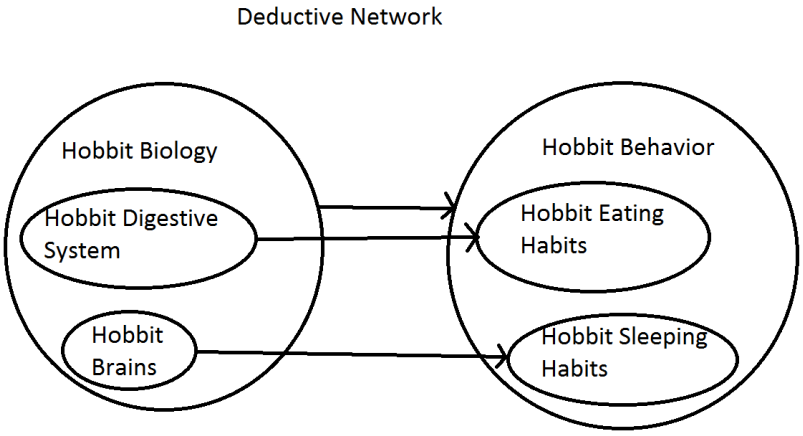
<d-link>
<d-premise>X4329</d-premise>
<d-conclusion>X4529</d-conclusion>
</d-link>
<d-link>
<d-premise>X4330</d-premise>
<d-conclusion>X4530</d-conclusion>
</d-link> <d-link>
<d-premise>X4331</d-premise>
<d-conclusion>X4531</d-conclusion>
</d-link>

Note that these links are defined to fit the contents of the article themselves. Meaning for example, the actual text of section on Hobbit sleeping habits in the second paper, refers to the section on Hobbit brains in the first paper. Not only that, but there's a flow of reasoning in the link. ie: understanding Hobbit brains is necessary to understand Hobbit sleeping habits. Who writes this file? The author of "Hobbit Behavior in Middle Earth" would, but the ID's might be assigned by some central authority that ensures uniqueness.

From this linkage file, along with the two definition files, we can construct a DN:

(F.2)

Figure 2:



The differences between Figures 1 and 2 should be noted. Figure 1 is simply a citation-based network. Meaning one article references the other. There's no implication of a flow of reasoning, or a deduction from the first article to the second. Figure 2 is a DN. There's an implication that there's a flow of reasoning in the direction of the arrows. There's also an implication that each node is a cohesive chunk of knowledge.

Also, note that the ID of each chunk of knowledge can be used to determine the article it corresponds to. In the background, there will be a database of records of "knowledge chunks" and each of these records will have a field corresponding to article-id or something of the sort. So although Figure 2 doesn't display article information, that information can be pulled, or perhaps displayed on a mouse-over.

7 How would DNs work in a research context?

For simplicity, assume all journal papers are in some central database, with definition and linkage files as described above. A computer can display all DNs based on these journal papers. A user can zoom into a part of the network he/she finds interesting. Note that Figure 2 doesn't display the

knowledge within each node. For that perhaps, a user will click on a node. If the journal is public, the knowledge will be displayed. Otherwise perhaps the user will be prompted to buy the journal article, after which he can see the knowledge in the node.

Is the DN useful even though it doesn't display the knowledge within each node? Yes. The title (if well chosen) of the node may reveal sufficient information for a user to know if the knowledge in the node is of interest. Also, perhaps a user is in search of the most important articles that a specific article references. He could use a measure like the number of deductive links to the article of interest, to determine if a reference is important or not. The DN alone can give information about the chain of reasoning involved. Perhaps a user is only interested in a certain aspect of a paper, and only wants references that deal with that aspect. A DN could assist with this.

Why not just read the articles to find the chains of reasoning?

To save time. Reading an article, and finding the references you're interested in is hard enough. But then you may have to do that with the reference article again, and so on multiple times. Having a visual display of a network will greatly save time and effort.

8 Can computers use DN's to gain knowledge?

DNs are not perfectly rigorous chains of logic. They're informal tools for use by humans, to follow chains of reasoning. Nevertheless they possess more structure than raw paragraph based texts. They also capture the deductive structure of knowledge, something that is missing in purely semantic networks. So we would contend that the step from DN's to computerized knowledge is far shorter than the step from natural human language to computerized knowledge.

References

[1] Chris Belter. Visualizing networks of scientific research. URL `http://www.infotoday.com/online/may12/Belter-Visualizing-Networks-of-Scientific-Research.shtml`.

[2] Damien George and Rob Knegjens. Paperscape. URL http://paperscape.org/.

[3] Ameet Sharma. Proposal for developing an xml system in scientific papers and articles, to capture the deductive structure of knowledge. URL http://vixra.org/abs/1701.0668.