

Capturing the user's reading context for tailoring summaries

Cécile Paris and Stephen Wan

CSIRO ICT Centre, Locked Bag 17,
North Ryde, NSW 1670, Australia
{Cecile.Paris, Stephen.Wan}@csiro.au

Abstract. The web has become a major source of information to learn about a topic. With the continuous growth of information and its high connectivity, it is hard to follow only the links that are relevant and not to get lost in hyperspace. Our aim is to support people who read documents in a highly connected information space, helping them remain on focus. Our contextually-aware in-browser text summarisation tool, IBES, does this by capturing users' current interests and providing users with contextualised summaries of linked documents, to help them decide whether the link is worth following.

Keywords: user's interest; tailored summaries; browsing support tool

1 Introduction

The web has become a major source of information to learn about a topic. With the continuous growth of information and its high connectivity, it is hard to follow only the relevant links and remain focused. While reading a document, people often encounter a promising link, which they decide to follow, only to discover after a quick browse that the document is not relevant to their current needs. Often, to avoid losing their focus, people open the linked document into a new tabbed window to which they return later. In such cases, users usually have many tabs opened. When they finally get to the linked documents to read them, they sometimes wonder why they opened these documents in the first place.

Our aim is to support people who read documents in a highly connected information space. In particular, we want to provide them with support to remain on focus. Our contextually-aware in-browser text summarisation tool, IBES, does this by capturing users' current interests and enabling users to obtain summaries of linked documents to help them decide whether the link is worth following. Then, when they follow a link, the system reminds them of their interest at the time they opened the document. Importantly, the summaries IBES generates are not generic, but rather are tailored to the user's current interests.

It is difficult in the general case and without *a priori* interactions for a system to know a user's interest. We hypothesise that what a user is currently reading reflects his or her immediate interest. We exploit this snapshot of the user's current interest to produce tailored summaries of related documents. We also provide an interface

enabling them to remain focused on what they are reading and to remember how they reached a specific document.

2 The In-Browser Elaborative Summariser (IBES) System

2.1 Overview

The IBES system is an internet browser¹ plug-in designed to support users browse through a large amount of information in order to learn about a topic. Based on the hypothesis that the user's reading context is a convenient, even if approximate, snapshot of the user's current interest, IBES obtains and captures this information through simple and efficient methods. It essentially notes the current page and the specific sentence of interest through a mouse-over movement: when the user moves the mouse over an anchor text link, this indicates to IBES the specific interest. This information is then exploited to generate a summary of the linked document.

The summary is generated using extraction-based summarisation techniques (cf: [1]) and is tailored to the current reading context. This summary thus acts as a preview of the document in relationship with the current document. It is provided to the user within their reading context, in a popup window. This enables the user to stay in focus. The user can then decide, based on this preview summary, whether the link is worth following or not.

The IBES System is illustrated in Fig. 1 below, using a Wikipedia text. It can be characterised as follows:

- *User Need*: Tell me more about the sentence that I have just read (the linking sentence), using content from the linked document.
- *Possible User Tasks*: Verify the statement just read; Learn more about that proposition; Decide if the linked document is worth reading.
- *Interaction*: The user moves the mouse over the hyperlink. This sets the linking sentence as the user's interest. This is passed to the summariser.
- *System Output*: IBES pops up a window which provides a preview for the linked page. It contains the first sentence of the linked document and a dynamically created summary that is tailored to the current user interest.
- *Interaction*: Having read this preview, the user can go to the linked page (see Fig. 2) or simply close the popup window.

In the example of Fig. 1, the user is reading about Louis Pasteur and moves his mouse on the link for "microbiology" in the sentence: "*He is regarded as one of the three main founders of microbiology, together with Ferdinand Cohn and Robert Koch.*" IBES takes this sentence as the user's reading context and generates an extractive summary of the linked page, taking this context into account (the Summary in Context). Here, 6 sentences out of 78 were extracted from the linked page, and all relate to the founding of microbiology. The popup also includes the first sentence of

¹ We currently work with the Firefox browser.

Louis Pasteur

From Wikipedia, the free encyclopedia

Louis Pasteur (December 27, 1822 – September 28, 1895) was a French chemist and microbiologist born in Saint Cloud, Paris. He is best known for his remarkable breakthroughs in the causes and prevention of disease. His experiments supported the germ theory of disease, also reducing mortality from puerperal fever (childbed), created the first vaccine for rabies. He was best known to the general public for inventing a method to stop milk and wine from causing sickness - this process came to be called *Pasteurization*. He is regarded as one of the three main founders of microbiology, together with Ferdinand Cohn and Robert Koch. Pasteur also made many discoveries in the field of chemistry, most notably the molecular asymmetry of certain crystals.^[1] He is buried beneath the Institute Pasteur, a rare honour in France, where being buried in the cemetery is mandatory save for the fewer than 300 "Great Men" entombed in the Panthéon.

Contents [hide]

- 1 Early life and biography
- 2 Work on chirality and the polarization of light
- 3 Germ theory
- 4 Immunology and vaccination
- 5 Allegations of deception
- 6 Honours and final days
 - 6.1 Statements
- 7 See also
- 8 References
- 9 External links

Early life and biography

Louis Pasteur was born on December 27, 1822 in Dole in the Jura region of France and grew up in the town of Arbois.^[1] There he

Microbiology

First Sentence:

- Microbiology (from Greek, *mikros*, "small"; *bios*, "life"; and *-logia*) is the study of microorganisms, which are unicellular or cell-cluster microscopic organisms.

Summary in Context:

- [25] While Van Leeuwenhoek is often cited as the first microbiologist, the first recorded microbiological observation, that of the fruiting bodies of molds, was made earlier in 1665 by Robert Hooke.
- [26] The field of bacteriology (later a subdiscipline of microbiology) is generally considered to have been founded by Ferdinand Cohn (1828-1898), a botanist whose studies on algae and photosynthetic bacteria led him to describe several bacteria including *Bacillus* and *Beggiatoa*.
- [27] Cohn was also the first to formulate a scheme for the taxonomic classification of bacteria.
- [28] Louis Pasteur (1822-1895) and Robert Koch (1843-1910) were contemporaries of Cohn's and are often considered to be the founders of medical microbiology.
- [34] While Pasteur and Koch are often considered the founders of microbiology, their work did not accurately reflect the true diversity of the microbial world because of their exclusive focus on microorganisms having direct medical relevance.
- [35] It was not until the work of Martinus Beijerinck (1851-1931) and Sergei Winogradsky (1858-1953), the founders of general microbiology (an older term encompassing aspects of microbial physiology, diversity and ecology), that the true breadth of microbiology was revealed.

6 of 78 sentences

Fig. 1: A summary generated when moving the mouse over “microbiology”

the linked document, which, in Wikipedia, describes the main entity for the article.

Showing the summary as a popup window in the user's browser page enables readers to keep their focus on what they are reading. Presenting a summary tailored to the reading context enables them to be in a better position to decide whether to follow the link: they can tell from the contextualised summary whether the linked document is likely to be relevant to their current focus, or they might have obtained already the information they were seeking. IBES highlights the relevant words in the contextualised summary. The popup window also contains the link (not shown) so that readers can easily follow it. If the user decides to follow the link, IBES will remind users as to how they reached the new page, thus further supporting users to maintain their focus, and it will provide a link back to that previous reading context. This is illustrated in Fig. 2. Finally, IBES highlights on the linked document the sentences that were included in the Summary in Context for ease of navigation in the page (not shown).

Note that another link to the same document will have a different reading context, and thus a different summary will be generated, as seen in Fig. 3. The reading context is on the profession “microbiologist”, and the sentences extracted, this time, are more

about the description of the discipline. This summary and that of Fig. 1 only have 1 sentence in common.

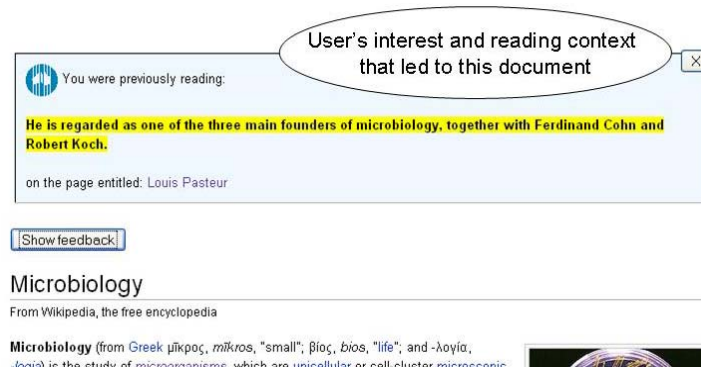


Fig. 2: In the linked document: IBES reminds the user of its original reading context

2.2 Capturing the user's interest

IBES considers that the current document represents the general user interest, and the linking sentence the specific one. It captures these two aspects. IBES currently performs only limited language processing (e.g., sentence segmentation). All sentences are represented using a vector space approach [2]. The simplicity of methods like the vector space approach is appropriate in contexts where the user needs to see and understand how the summaries are generated. The linking sentence is also recorded as such, so that IBES can use it to remind users of their original reading context. This method is simple and scalable. It does not require any information about the domain and does not need the identification of the user.

2.3 Generating the contextualised summary

When the user moves the mouse over a link, the IBES extension is triggered and provided with three pieces of information:

1. The linking page: the contents of the page being read;
2. The linking sentence: the text of the sentence in which the link is embedded; and
3. The linked page: the contents of the linked page to be summarised.

In extractive summarisation, a document is analysed to find its key words, assumed to represent what the document is about (cf: [1]). Representative sentences containing these words are then selected to form a generic summary of the document. In IBES, instead of choosing the sentences based on the key words of the document, the system chooses sentences that are related to the linking sentence. As mentioned, sentences are represented with vector space approaches. The cosine metric, a simple, scalable and fast method, is used to compare the vector space representation of each sentence in the linked page to the vector for the linking sentence.

Microbiologist

From Wikipedia, the free encyclopedia

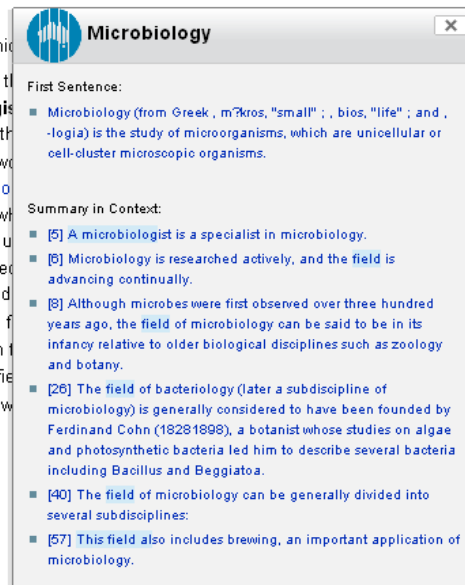
A **microbiologist** is a scientist who works in the field of **microbiology**. Most have a university degree in the subject.

Specialists in the broad field of microbiology include:

- **Bacteriologists**, who work in the field of bacteriology
- **Environmental microbiologists**, who study microbial processes in the environment
- **Food microbiologists**, who work on microorganisms that cause food spoilage
- **Industrial microbiologists**, who work on microorganisms that produce useful products
- **Medical microbiologists**, medical microbiologists specialize in the diagnosis and treatment of infectious diseases
- **Mycologists**, who work in the field of mycology
- **Protozoologists**, who work in the field of protozoology
- **Virologists**, who work in the field of virology
- **Microbial epidemiologists**, who study the spread of infectious diseases

Famous microbiologists include:

- Anton de Bary
- Martinus Beijerinck
- Jules Bordet
- David Bruce



The screenshot shows a window titled "Microbiology" with a close button (X) in the top right corner. The window contains the following text:

Microbiology

First Sentence:

- Microbiology (from Greek, *mikros*, "small"; *bios*, "life"; and *-logia*) is the study of microorganisms, which are unicellular or cell-cluster microscopic organisms.

Summary in Context:

- [5] A microbiologist is a specialist in microbiology.
- [6] Microbiology is researched actively, and the field is advancing continually.
- [8] Although microbes were first observed over three hundred years ago, the field of microbiology can be said to be in its infancy relative to older biological disciplines such as zoology and botany.
- [28] The field of bacteriology (later a subdiscipline of microbiology) is generally considered to have been founded by Ferdinand Cohn (1828-1898), a botanist whose studies on algae and photosynthetic bacteria led him to describe several bacteria including *Bacillus* and *Beggiatoa*.
- [40] The field of microbiology can be generally divided into several subdisciplines:
- [57] This field also includes brewing, an important application of microbiology.

Fig. 3: Another summary of the page "Microbiology", but from a different reading context.

3 Related Work

Our work falls under "user-focused" (or topic-focused, or query-focused) summarisation (cf: [3]; also, e.g., [4]) in which a summary takes into account some representation of the user's interests, typically as indicated in a profile or from a question/query. In IBES, users do not have to issue queries, and there is no need for an *a priori* profile. A user's current reading is taken to be his or her current interests. Also, our summaries are not generated in a search context, but as support to the task of browsing through cited documents while reading a specific article.

Amitay and Paris [5] exploited link text for summarisation, recycling human-authored descriptions of links from anchor text to generate web-pages summaries. In our work, we use the anchor text to provide a context for the summary.

Other researchers have studied aspects of graph theory applied to summarisation, e.g., [6, 7], although they do not focus on a live reading context as we do.

Other work has addressed the issue of capturing the user's interest for Web personalisation, e.g., [8, 9]. Our work is more concerned with exploiting user's interests to tailor summaries of related documents and thus can be seen as complementary. We were interested, however, in using as simple a technique as possible to capture the user's interest.

4 Discussion

IBES provides a summary of a document tailored to the user's current reading context within a browser, enabling a reader to get an overview of a linked document without losing his or her focus. Currently, the reading context is taken to be the sentence in which the link occurs. We have done experiments with the size of this window to model the user's interest, but our results so far are inconclusive. We will continue to explore this issue. We will also investigate tracking users' interests over time.

IBES currently works on Wikipedia, but its underlying modules are generic. There are only two features of IBES that are specific to Wikipedia: 1. the inclusion in the preview of the first sentence of the document. We thought that this overall description of the page would be useful in a preview; 2. the module that strips off navigation panels/etc. to provide access to the text proper. We are studying algorithms that would work on arbitrary web pages. Note that we have applied the work to a different data set: scientific articles (linked through citations).

We are designing an end-user evaluation to determine the utility of our context-sensitive summaries. In future work, we intend to explore additional summarisation strategies and the applicability of the tool for other types of document.

Acknowledgments. We would like to thank Julien Blondeau for his work on the system and Nathalie Colineau for her comments on an earlier draft of this paper.

References

1. Mani, I.: Automatic Summarization. John Benjamins Publishing Company, Amsterdam/Philadelphia (2001)
2. Salton, G., McGill, M. J.: Introduction to modern information retrieval. McGraw-Hill (1983)
3. Firmin, T., Chrzanowski, M. J.: An Evaluation of Automatic Text Summarization Systems. In: Manni, I., Maybury, M.T. (ed.): Advances in Automatic Text Summarization. MIT Press, Cambridge, Massachusetts (1999) 325-336
4. Berkovsky, S., Baldwin, T., Zukerman, I.: Aspect-Based Personalized Text Summarization. In: 5th International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems (2008) 267-270
5. Amitay, E., Paris C.: Automatically summarising web sites: is there a way around it? In: 9th Int'l Conf. on Information and Knowledge Management (2000)
6. Teufel, S., Moens, M.: Summarizing scientific articles: experiments with relevance and rhetorical status. Computational Linguistics, MIT Press 28 (2002) 409 - 445
7. Mani, I., Bloedorn, E.: Summarizing similarities and differences among related documents. Information Retrieval 1 (2000)
8. Sia, K.C., Zhu, S., Chi, Y., Hino, K., Tsen, B: Capturing User Interests by Both Exploitation and Exploration. In: the 2007 International Conference on User Modelling (2007) 334-339
9. Zhang, H., Song, Y., Song, H.: Construction of Ontology-Based User Model for Web Personalisation. In: the 2007 International Conference on User Modelling (2007) 67-76