Embedded Definition Interfaces

Steve Balensiefer and Tyler Robison
University of Washington
{alaska, trobison}@cs.washington.edu

Abstract

In this work we explore and evaluate several interfaces for augmenting text with the definitions of certain words; here we use a running example of embedding definitions of obscure and technical words in medical text. We step through our design process, showing the prototyping and user evaluation we went through, and explain our reasoning and decisions as we go. Our results suggest that, among other things, there is significant interest in such interfaces provided that they do not fundamentally change, but merely augment, the reading process.

1. Introduction

Despite living in this digital age, and despite the myriad ways in which our everyday tasks have been improved or otherwise altered by technology, the way in which we read text is fundamentally unchanged from the way people have always read; reading tends to be a largely static affair, with little user interaction beyond flipping a page or clicking a 'next' button. HTML documents are an exception, and their links change the process of reading from a primarily linear process to one with options and possible user interaction.

Yet such advances, the digital medium is not being fully exploited. Who, in the process of reading a digital document, has not been forced to look up an obscure or technical word? One standard response here is to look up the word in one of many online dictionaries, but this interferes with the reading process, and currently requires copying the word, switching applications, etc. Yet why should the user be forced to supply the connection in this otherwise digital system?

In this work we investigate ways to reveal such definitions to users without getting in the way of the reading process while still serving as a useful reading aid; we are interested in tools that let the reader use the document in exactly the same way they always have, but with the additional functionality of having accessible definitions. We focus on the way in which these definitions are shown, and the ways in which the user interacts with them, and do not consider what definitions are used or how they get there; we assume that the document's author provides relevant and useful definitions for the purposes of this work. That said, we provide some initial thoughts on how this might be done in section 9.

This work was developed with the context of reading technical documents outside of one's field in mind, and all of the text we used in our surveys and studies is medical text found online; throughout all of our experiments we look at using our embedded definitions to allow laypeople to read difficult medical text. That said, we do not evaluate the effects of these interfaces on improving the understanding of medical text, we only evaluated them in terms of usability. While we have only studied these interfaces in the context of reading medical text, we believe that they could be useful in reading other types of material, based on comments of our study participants.

Our decision to focus on reading medical text was motivated by the example of giving laypeople access to otherwise difficult medical text, though, as mentioned above, we believe our findings are not restricted to this example. Throughout this project we used samples of text in our surveys and observational studies, and these samples were taken from online medical research journals; we wanted to use especially difficult text for our studies, to necessitate looking up definitions. The samples selected were the sorts of topics a layperson may come across in reading about a condition online, but were somewhat more technical than what one might find in many cases.

We proceed in the next section by looking at some related work in the area, some of which strongly guided our initial thoughts and prototypes. In section 3 we go over our initial ideas, and describe the paper survey conducted to gauge interest in the area. In this section we also introduce our 4 evaluation criteria; these are the usability characteristics of interest in our interfaces, and which were used by our study participants in their evaluations. In sections 4, 5 and 6 we describe how we created, evaluated and

expanded prototypes in an online survey and 2 observational studies, and the results obtained along the way. We list our conclusions and describe future work in sections 7 and 8. In section 9 we briefly discuss issues that were beyond the scope of this work, but which would need to be resolved if a such system were to be deployed.

2. Related Work

Our goal of providing embedded definitions is an attempt to realize one of the many possible uses for "hypertext". In 1965 Nelson visualized a system that would allow a body of information to be connected in a way that could never be statically shown on paper [3]. The area of hypermedia itself developed in the late 70's and grew in various ways in the 80's, with such tools as Hyperties [2].

The Fluid Documents Project [6] work during the late 90s at the Palo Alto Research Center (PARC) investigated some of the issues in providing interactive content along with static documents. In that work, they were investigating ways to provide additional information to user about the content of links provided in web-pages. They appropriated the idea of a *gloss*, a short description between lines or in the margin of textbook. The initial work focused on glosses as a mechanism to assist in Web Navigation [5], while a later paper examined the effects of gloss presentation on reading [6]. This work guided our early thinking in numerous ways, and our initial prototypes use some of the same ideas found in these fluid documents. Our work differs in that we explore some areas untouched by the fluid documents, such as multilevel definitions (described later), and the effects of different highlighting techniques, and because we approach evaluation from a usability perspective based around 4 evaluation criteria (also described later).

The Adobe Acrobat Reader [4] provides the ability to lookup definitions by highlighting a word, right-clicking, and selecting an option from the menu. "Dictionary Search" for the Firefox browser provides similar functionality [7]. The actual mechanism is to submit a query to an online dictionary in the default web browser. This allows access to a larger range of definitions that our system, but requires the user to wait for the page to load. There is also no visual notification that this feature is available. Our design decision to include definitions in the actual text removes the page loading delays, and none of our interfaces require menu selections; such menus seem to place an unnecessary barrier between the user and the document.

Another relevant work is a commercial product called IntelliTXT™ from Vibrant Media [1]. This is a system that inserts advertising in the form of small pop-up windows (like a tool-tip) when users move their mouse over special keywords. We try a similar interaction technique for providing definitions, in addition to others, and with evaluation.

Finally, the web itself is full of hypertext, and various interfaces such as tool-tips are fairly common online. Our work tries to shed light on which interfaces are better than others, and why, in addition to exploring other areas such as multilevel definitions.

3. Initial Stages

3.1. Initial Ideas

At the beginning of this project we wanted to iteratively design and evaluate an interface to integrate definitions in text, but as we looked at the related work in the field, we saw that there were already a number of good ideas out there, and so decided to investigate and evaluate some of the more common, and appealing, themes, such as those seen in the Fluid Documents work.

3.2. Paper Survey

Before we committed ourselves to the project, we wanted to ensure that something like this would be useful. So we conducted a paper survey in which the participants were given a paragraph of technical medical text and asked to read and try to understand it. Next they were asked how much they understood, were asked about how they normally react to unknown words, and finally how useful built-in definitions would be to them. The purpose of all of this was to gauge interest in embedded definition systems; the reading of the medical text was to give them exposure to the types of situations we were looking to address.

We gave the survey to a total of four people, all of whom were graduate students in the Computer Science department. All participants said that it would be useful to have definitions available, assuming the system was not intrusive, and that the definitions were applicable in the context of the reading, and at the proper level of detail. With regards to a definition system one participant wrote:

> "The unfamiliar words stop me already, so being able to 'summon' a definition wouldn't take much more time."

In summary, the initial study suggested that people would use a system that provided definitions for technical words, if it satisfied certain conditions. The particular conditions they mentioned were 1) that the definitions be at the appropriate level (not too technical, for instance), and 2) that the system be non-intrusive. We used these comments to help determine the evaluation criteria we used in later studies. These evaluation criteria are explained later in this section.

In this survey, and in the next survey and both observational studies of this work, all of our participants were Computer Science graduate students. As such, we can't say with certainty that our results generalize; computer scientists work with technology on a regular basis and are likely to be familiar with interfaces similar to those used throughout our project, and this may bias the results. That said, the aspects of the interfaces used here should be familiar to many web-surfers, and we see no reason to believe that computer scientists should have drastically different preferences in such interfaces than other people. Therefore, while we acknowledge that we haven't shown that the results generalize, we believe that they are not unique to computer scientists, and could well extend to broader groups.

3.3. Evaluation Criteria

From the onset we had in mind a list of characteristics that we considered important for a good interface. We had arrived at this list through discussion with each other, and with our own experiences, both good and bad. There was also some influence from the initial surveys conducted; one participant stated that a definition-providing interface would be useful as long as it was not 'invasive', an idea not dissimilar from our notion of unobtrusiveness below. The purpose of laying out these criteria was first to focus our thinking on what would be useful in an interface, but more importantly to help us evaluate the interfaces in later experiments.

Interaction Clarity: How easy is it to understand the interface; there shouldn't be any guesswork as to what can be interacted with, how to interact with it, and what happens when you do. It should respond to user input as your intuition would suggest. Anyone who uses a web-browser should be able to use the interface with a minimal learning period.

Unobtrusiveness: The traditional way of reading the text shouldn't be changed; if you don't want to use the new features, you should still be able to read the text in the same way as you normally do, without any extra distractions or hindrances. The interface should not demand anything of you.

Flow Preservation: When using the new features (getting a word's definition), the 'flow' of reading the text shouldn't be broken; that is, you should be able to access a definition without losing your place in the text, or losing your train of thought.

Ease of Use/ Effort Required: The amount of effort required to use a new feature (get a word's definition, for instance) should be small; having to pull out a dictionary, turn to the right page, and find the entry on that page is more work than having it given to you on a slip of paper, for instance.

Although we believed these criteria adequately covered the important issues of an embedded definition interface, we wanted to get the opinions of users in regards to these criteria, and did so in our online survey.

4. Online Survey

4.1. Purpose

The paper survey had only served to show us that there was interest in embedded definitions, but it had not given us the details of what might be important and useful. In order to get more information, we decided to build a variety of prototypes, and run them by users in an online survey. We chose an online survey in the hopes that it would reach more people; users could take it from the convenience of their own computer, and do so on their own time.

The survey was designed so that we would get feedback on each of the prototypes (described below), both in terms of how users rated them and in terms of comments, and so that we could ask a number of miscellaneous questions about user preferences. We chose to get feedback on numerous interfaces because we weren't sure which would be useful, and to try and eliminate personal bias, but also because we wanted to get at the underlying concepts of these interfaces; what were the general themes that one finds in all of the popular interfaces. Also held up to scrutiny here were our evaluation criteria; users were asked which of the 4 was the most important, and were asked if there were any additional criteria they considered important.

4.2. Interface Prototypes

Many of the prototypes we designed were patterned after some of the ideas in the Fluid Documents work; none of them are exact replicas, but there are many common concepts. We had originally intended to design our own prototypes, but as there were already numerous ideas out there, many of which struck us as good ideas, we opted for variations on these existing ideas. These variations were mainly changes we thought would make the interface more useful, or to make it fit in more naturally in the context of a web-browser. All prototypes were implemented in a web-browser using JavaScript.

Below we present the four basic categories of embedded definition interfaces; each takes a different approach in supplying the user with related definitions. In two of the four categories we designed several interfaces that were variations on the common theme of that category; the other two categories had only one interface each.

4.2.1. Glossary

Definitions are shown in a glossary frame separate from the main text, in the order they appear in the text. There were 2 dimensions of variation on theme; first, how the user interacts with the glossary, and second, where the glossary is located (top, bottom, left or right). In terms of how to interact with it, there are three possibilities: first, if the user clicks on a defined word in the main text then the glossary scrolls to that word and highlights it in red. The second is similar, except that the user runs the mouse over the word to make the glossary focus on the definition and turn red. In the final variation, there is no connection between the main text and glossary; they are both simple html pages without any special interactions (definitions are never highlighted).

So there are a total of 12 possibilities for this category. In this survey we only used 5 of them: the four combinations of having the glossary on the bottom or right, and interacting via clicking or mouse-over, and the final interface used had the glossary on the right, with no special interaction. Figure A shows a screen-shot of a right glossary with mouse-over interactions. This category shares some similarities with the Fluid Documents concept of 'fluid margin', though there are several substantial differences, the most prominent being that the glossary definitions are visible at all times, while the margin gloss is only visible when a word is selected.

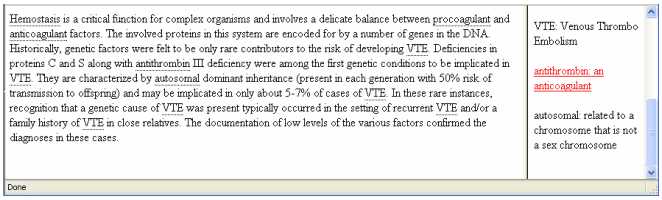


Figure A: The cursor is moved over the word 'antithrombin' in the right glossary interface, highlighting its definition in the glossary

4.2.2. Tool-tips

In this interface, when the user runs the mouse over a defined word a small box appears near the word containing the definition, possibly over existing text; this is the same idea of tool-tips as seen in various web-sites and application interfaces. When the mouse cursor is moved off of the word, the box disappears. We did not evaluate alternative versions of this, such as triggering the tool-tip by clicking instead of running the mouse over. A screen-shot of the tool-tips interface is shown in Figure B. This category shares similarities with the Fluid Documents concept of 'fluid overlay', with some slight differences.

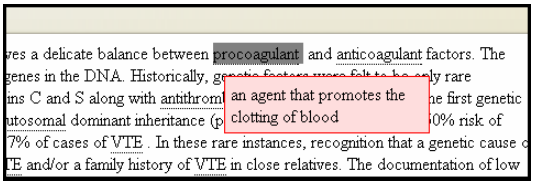


Figure B: The cursor is moved over the word 'procoagulant' in the tool-tips interface

4.2.3. Expand in Place

In this interface, definitions of words are inserted in the text after that word, pushing existing text off to the side, and these definitions are drawn in such a way as to stand out from the regular text; an example is shown in Figure C. There were three interfaces for this category; one in which the definition is triggered by left-clicking the word with the mouse, one in which the definition is triggered by moving the mouse over the word, and one in which all words are expanded all the time; the user cannot hide the definitions. In the first two, the user can hide a visible definition by clicking or running the mouse over the word, respectively. Alternatively, in the click triggered version, the user can leave several definitions expanded at once if they choose. This category shares some

similarities with the Fluid Document notion of 'fluid inline', with the main difference being that in our interface the text is inserted in the standard lines of the text, whereas the 'fluid inline' interface inserts the text below the word, in the shape of a tool-tip, but with the text pushed out of the way.

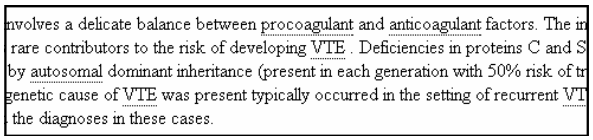
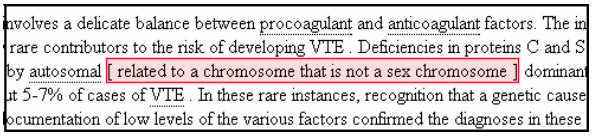


Figure C: Expand in place interface, before the word 'autosomal' is clicked



After 'autosomal' is clicked

4.2.4. Links

Each word in the main text is a link to a separate page that contains only the definition, so the user can interact with it exactly as they would a regular web-site. We did not use any alternative versions of this interface.

So there were a total of 17 interfaces, but only 10 were used in our survey, namely: 1) expand in place on click, 2) expand in place on mouse-over, 3) all expanded in place, 4-7) all four combinations of bottom/right glossary interacting via click & mouse-over, 8) right glossary without interaction, 9) tool-tips, and 10) links.

The only category in which we did not use all possible interfaces was #2, the glossary category. The reason is that many of these are very similar (glossary on the left versus the right), and we did not want to hit our evaluators with a legion of similar interfaces; while we no doubt could have collected some good comparison data, participants would be less willing to complete the survey if they had to look at 8 interfaces instead of 4, and many of them were redundant.

4.3. Survey Layout

Of these 10 interfaces, each user was asked to evaluate 4; participants following the link we sent them were randomly redirected to 1 of 4 versions of the survey. Below we list the interfaces used in each version.

Version 1: Expand in place on click, expand in place on mouse-over, links, and right glossary on click

Version 2: Expand in place on click, expand in place on mouse-over, right glossary on mouse-over, and the static glossary

Version 3: Expand in place on mouse-over, all expanded, tool-tips, and bottom glossary on click

Version 4: Expand in place on click, tool-tips, bottom glossary on mouse-over, and the static glossary

The tool-tips interface did not run correctly on the Microsoft Internet Explore browser, so users of that browser were randomly assigned to versions 1 or 2 (neither of which included the tool-tip interface); all other browsers were randomly directed to one of the 4 versions. Apart from the interfaces used, all survey versions were identical.

The version of the survey began with a statement explaining our purpose, followed by a list of the 4 evaluation criteria, along with a short description of each. In each of the versions of the survey, for each interface, the users were asked to read a paragraph of medical text (the same for each interface) and rate it in terms of the 4 evaluation criteria, on a scale of 1 to 5, with 1 being 'very poor', and 5 being 'very good'. For each interface they were also asked to rate it, on a scale of 1 to 5, on its overall usability. We were originally uncertain about using the same text for each interface evaluation, but since the participants were being asked to consciously evaluate the interfaces, we decided that it would be alright.

In all versions of the survey, after evaluating all four interfaces, the user was asked to answer a number of miscellaneous questions common to all versions of the survey. The first couple asked the participant which of the 4 evaluation criteria was the most & least important in designing a good interface, in their opinion. A number of other questions followed, asking about use of these interfaces in specific situations; we won't list them here in the interest of space. The final question asked whether multi-leveled definitions (initially getting a brief phrase definition, which, if the user wished, could be expanded out to a sentence definition, which could expand to a paragraph) would be useful (answers of 'Not at all', 'Somewhat', or 'Very'). At the end of the survey was a field for general comments.

4.4. Results

A total of 12 people participated in our survey, all of them graduate students in Computer Science. Their division between the different versions is as follows:

Version 1: 3
Version 2: 3
Version 3: 4
Version 4: 2

It would be difficult, and dangerous, to conclude much from the combined results of the 4 survey versions; the interfaces used in each were different, and while the users weren't asked to compare them with one another, there is likely to be a bias nonetheless. Seeing an interface with terrible flow-preservation may really bother them, and bias them toward saying that flow-preservation is the most important, for instance.

We can look at some statements suggested by the data, however, as long as we only regard them as indications. First of all, the only evaluation criteria that were selected as the most important were unobtrusiveness (selected by 4 participants), and flow preservation (selected by the other 8); neither interaction clarity nor ease of use were ever selected. This is interesting, since these two are fairly similar; unobtrusiveness is maintained if the reading of the document is unchanged when the definition features are not used, and flow preservation is maintained if the reading of the document is not disrupted when the definition features are used. Both require that the reading task is not fundamentally changed, that the interface does not impose any additional demands or distractions upon the user, and so it seems significant that they were selected as the most important.

Another interesting result of the survey was the list of the top 3 interfaces, according to their usability scores as rated by the participants. These three were, in order, the tool-tips, expand in place on click, and right glossary on mouse-over interfaces. What is interesting about this list is that each of these interfaces is quite different from the others; we don't see both expand in place options, or multiple glossary options, for instance. This suggests to us that different users have different preferences, that there is no one type of interface that is perfect for everyone. Of course, we did only have a single tool-tips interface, so it would be irresponsible to draw a conclusion from this data alone, yet the results of our observational study, described below, support this idea.

In the comments field of the survey we received numerous complaints and compliments on the interfaces, as well as suggestions of ways to improve them, and comments on already existing systems. What we did not see in the comments were any new evaluation criteria that we had missed; the text above the comment field asked specifically for such comments. The interpretation we drew from this was that our evaluation criteria did a fair job of expressing peoples' concerns; they encompassed the main aspects that users wanted in such an interface.

Finally, we saw that there was significant interest in multilevel definitions; only 1 of the 12 participants answered that such definitions would not be useful. From this show of interest, we modified our prototypes with multilevel definition capabilities for the next round of evaluations.

5. Multilevel Definitions Observational Study

The online survey gave us numerical data, but not the underlying thought processes that generated it; we saw what people liked and thought was important, but not why. So we decided to do an observational study to get at this issue of why; watching the participant as they used the interfaces, and listening to their comments, would allow us to get a more in-depth perspective. We decided to alter our prototypes with multilevel definitions, and explore this issue as well, since there was significant interest expressed in our online survey.

5.1. Multilevel Definitions

By multilevel definitions we mean that when the user requests a definition (say by moving the mouse over the word, in the tool-tips interface) a brief definition, a short phrase, is given. The user can then request more information (in a manner dependent on the particular interface), and the definition shown then expands into a more detailed version. So if the initial definition doesn't supply enough detail, or the user is just curious, they can expand the definition further to get more information. The idea was originally suggested to us by one response in our initial survey, and the results of our online survey indicated that our participants found this idea interesting, and believed that it would be useful.

There are numerous ways these multilevel definitions could be approached, and we decided to opt for a system in which the initial definition is a phrase, the next definition is a sentence or two, and the third definition is a few sentences; none of our words have more than 3 levels of definitions. We also did not impose any rules or

structure on the different levels of definitions, though one could conceive of doing so.

5.1. Multilevel Definition Prototypes

As was previously mentioned, we selected the top 3 interfaces of the online survey (as was determined by the usability ratings), and added multilevel definition capabilities to each. The way in which the user triggers another level of the definition varies from interface to interface. Only the top 3 were chosen as we wanted to perform a lengthier study on each of the interfaces, and so had to reduce the number of interfaces used to keep the study length at a manageable level.

5.1.1. Tool-tips on Mouse-over

This interface was extended from the tool-tips interface described in section 4.2.2. In this interface, a small window pops up below and slightly right of the defined word when the user moves her mouse cursor over the word (See Figure D). We extended this for multi-level definitions by letting the user click the keyword to expand the current definition. We also added text at the bottom of the tool-tip with the words "(click original word to expand)" if there was additional information to show. We did not allow users to contract the definitions in this interface, in part because the tool-tip definitions are only visible when the mouse is over the word, and so the user doesn't have to see them if they don't want to.

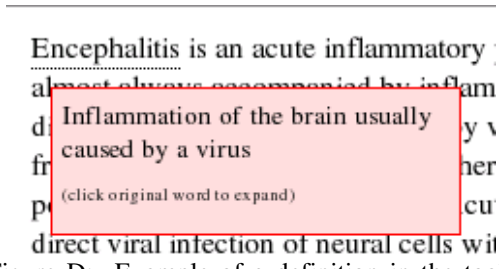


Figure D: Example of a definition in the tool-tips on mouse-over interface, with multiple definition levels.

5.1.2. Expand in Place When Clicked

This interface inserts the definition text into the document flow after the defined word. To support multiple level definitions we added two buttons designated "[expand]" and "[contract]" at the end of the definition text. Clicking the expand button accesses the next definition level, while contract returns to the previous definition level. If the definition is at the most basic level then clicking contract hides the definition. Similarly, clicking the original word hides the entire definition regardless of level. Definitions hidden this way are reset to the short version of the definition if the word is clicked again.

5.1.3. Right Glossary on Mouse-over

As in the expand in place interface, "[expand]" and "[contract]" buttons are added to the end of each definition in the glossary. The user can also expand a definition further by clicking on the word in the main text; they don't have to scroll over to the glossary window to do so. In order to contract, however, the user must scroll over to the glossary window and click on the contract button. See Figure E.

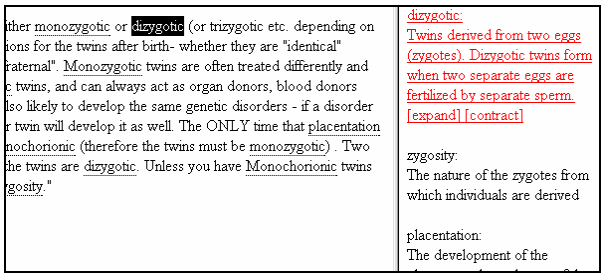


Figure E: Right glossary on mouse-over interface with multilevel definitions

5.2. Study Design

First they were read a brief description regarding embedded definitions, what they were about to do, and were asked to think aloud as they read through the text. They were informed that they were free to leave at anytime, and that their names would not be revealed in our use of the data. Next they were given a sheet listing the 4 evaluation criteria, and they were asked to read through it and ask any clarifying questions.

Each of the 3 interfaces used a different body of text, unlike the online survey, all of roughly the same length with roughly the same frequency of defined words (about 2 per sentence). And also unlike the online survey, each participant saw the same interfaces (though in a different order). Upon being given an interface with its associated body of text, the participant was asked to read through it with the aim of understanding it; that is, they were not just asked to play with the interface. They were encouraged to think aloud, and say what they were doing and why doing the exercise. After finishing the paragraph, the user was given an evaluation sheet and asked to fill it out, commenting on their decisions as they went. They were also encouraged to play with the

interface further if desired. The evaluation sheet was similar to that of the online survey; the participants were asked to rate each interface in terms of the 4 evaluation criteria, and in terms of both 'how well they liked it' and 'how likely they would be to use it', separately (different from the online survey).

After evaluating the interfaces, they were given a final sheet of questions, asking them to select the most and least important of the 4 evaluation criteria, along with a couple of miscellaneous questions. They were asked to comment on their choices on all evaluation sheets and the final questions, and were allowed to modify their answers to any of the evaluation sheets at any time during the study (to change an old answer in light of having seen a new interface, for instance).

Before running the actual study, we performed a pilot study to iron out some potential problems. Through it we fixed numerous small wording problems, and added some clarifications. The largest change was that we decided to use a different body of medical text in each interface; in the original version each interface used the same text, but it made later readings too artificial, and so each interface was given its own text.

5.3. Results

We conducted the observational study with a total of 4 participants, all of whom were graduate students in the Computer Science department. There were an equal number of men and women.

Although participants were asked to rate each of the interfaces in terms of the evaluation criteria, in terms of how much they liked it and how likely they would be to use it, the real value of these studies was in the comments and observations that resulted; from these studies we received the kind of helpful details that were lacking in the online survey. In the process of rating each interface in terms of each of the evaluation criteria and answering the other questions asked, participants provided a wealth of comments on their preferences, dislikes, annoyances and suggestions. It is not within the scope of this paper to sort through all of this data, and so we present some of the most relevant results.

One of the first findings that leapt out at us was one suggested first by the online survey, that different people frequently have very different opinions. What annoys one person in an interface is perfectly acceptable to another, and what is a useful feature to one may be an unnecessary complication for another. For instance, participant #3 thought it was unnecessary and distracting for a defined word to be expandable several times in the same paragraph; he would prefer that only the first instance of a word in the paragraph has an attached definition. Yet participant #4 found repeated definitions very useful; he used their definitions as reminders of what he had read before. There were several other instances like this, where one person's opinion of what should be done was directly opposed to another's opinion.

In these studies, 3 of the 4 participants (all except for participant #1) preferred the tool-tips interface over both the expand in place on click and right glossary on mouse-over (the other preferred the expand in place interface); this is taken from both their comments and their numerical data. This makes sense in light of our online survey results, in which the tool-tips interface was rated number 1. Participants provided many comments on why; in short, because using a tool-tip doesn't disrupt the text or require you to move your gaze or mouse elsewhere. The dissenting participant valued the expand in place interface's ability to have multiple definitions out at once, something that wasn't possible with the tool-tips, and expanding in place wasn't as distracting to her as it was to other participants.

No one selected the glossary as their preferred interface; some disliked it more than others, but all 4 of the participants commented on how moving your gaze from the main text to the glossary disrupts the regular flow of reading.

The same 3 participants that preferred the tool-tips interface selected flow preservation as the most important evaluation criterion, while participant #1, who selected expand in place, chose unobtrusiveness. This selection of evaluation criteria is similar to what we saw in the online survey; there 2/3 of the participants selected flow preservation, and the remaining 1/3 selected unobtrusiveness as the most important. It is appropriate that the participants who chose flow preservation as the most important criterion would choose tool-tips as their preferred interface, as it tends to disrupt the reading less, and that the participant who chose unobtrusiveness also chose the expand in place interface, as it is less obtrusive than the tool-tips, though with worse flow preservation.

Another strong pattern seen in the comments and ratings is one also seen in the online survey, that users want the new features to be added in

such a way that the basic process of reading the text is not greatly changed. We see this in the choice of the most important evaluation criterion (with the choices being flow preservation and unobtrusiveness), and in participant comments; there were numerous comments to the effect that anything drawing one's attention from the current point in the text is undesirable. Definitions should be available, but only present when called for, and should not draw the user's attention from the text.

Overall, the participants found the multilevel definitions at least somewhat useful, though there were a couple of comments saying that they would be more useful if these leveled definitions conformed to a more strict structure; for instance, the original definition could give the basic idea, the next level could describe it in the context of the document, and third level could provide an example, if applicable. Participants suggested they would find they more useful if they knew before hand what sorts of information they would get on later levels.

6. Highlighting Observational Study

In the previous observational study we had heard comments on how words emphasized through use of a dotted underline were fairly non-distracting, and would not be confused with links in web-pages. Through some discussion with others, and amongst ourselves, we decided to conduct a short study on the different ways to highlight interactive words in an embedded definition interface.

We wanted to test different methods of when to highlight interactive words, and within each such method, what technique should be used to highlight (dotted underline, different color, italics, etc.) We were pleased with the results of the previous study, and so decided to use the same study technique here as well; namely, a think-aloud observational study.

6.1. Prototypes

We had 3 such methods in mind for when to highlight words; the first is one we had used in all previous interfaces, that is, if a word can be interacted with, it is always highlighted. We'll refer to it as the '*always highlighted*' method. The second is one in which nothing is highlighted by default, but when the user moves the mouse over a paragraph, all words within that paragraph that can be interacted with are highlighted; we'll refer to this as the '*paragraph highlighting*'. Finally, the third method is one in which the user can toggle whether words are highlighted or not; there is a button (in our tests it was in the web-page, but in a real implementation it would be a part of the browser) that toggles whether highlighting is shown or not. If the highlighting is turned off, words can still be interacted with (say, by running the mouse over a word to bring up a definition, if it is the tool-tips interface), but there is simply no highlighting to indicate this. We'll refer to this as '*toggled highlighting*'.

Each of these methods was implemented on top of the tool-tips interface used in our previous observational study; now we had 3 different versions to test, all of them using the tool-tips interface. In each, we wanted users to be able to select the means used to highlight interactive words (whether it was a dotted underline, red text, italics, etc.); to do this we had a series of buttons at the top, one for each possible highlighting technique. The user was allowed to dynamically change the technique of highlighting used while using the interface, so that they could compare different ways of highlighting and arrive at their favorite. The 7 different highlighting techniques are as follows: 1) bold 2) italics 3) inverse (black background, white letters) 4) red letters 5) dotted underline 6) larger font and 7) plain text (no different from non-interactive words). This button selection of highlighting was used in all 3 of the interfaces, and is shown in Figure F, a screen-shot of the 'always highlighted' interface.

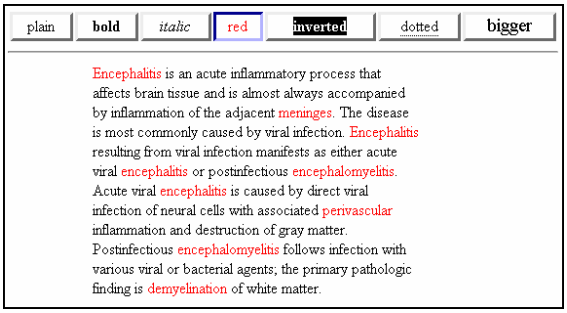


Figure F: A screen-shot of the 'always highlighted' interface, with the highlighting selecting buttons shown at the top.

6.2. Study Design

We began by describing our project of investigating embedded definitions in text to each participant, and the aim of the experiment to look at use of highlighting. We then informed the participants that they were free to leave at any time, and that their names would not be revealed in our data. The study consisting of showing them the 3 interfaces, in varied order,

each of which used the same text, and for each interface having them select the best highlighting technique (dotted underline, red, italics, etc.) After seeing all 3 interfaces, they were asked to select their favorite of the 3; which one they would like to use the most. Unlike the online survey and the previous observational study, there were no evaluation sheets. And, unlike the previous observational study, they were not asked to read and try to understand the text, only to choose the means of highlighting; hence we decided that it would be alright to use the same text for each interface. This text had about the same frequency of defined words as in previous studies; about 2 per sentence. We speculate that the preferred means of highlighting may be dependent on the frequency of the defined words, but have not tested this due to time constraints.

As in the previous observational study we ran a pilot study before conducting the actual tests, though in this case no substantial changes were needed.

6.3. Results

The study was conducted with a total of 3 participants; 2 men and 1 woman, all of whom were computer science graduate students. We had originally hoped to use at least 4 participants, but due to time constraints we were unable to do so.

Overall, when using the paragraph or toggled highlighting, participants did not like highlighting techniques that changed the shape of the text; it was jarring for all of them when the text moved around because the font was slightly larger than normal. This was the case for the bold, italics, inverted and bigger techniques. Furthermore, 2 of the 3 participants mentioned that italics and bold could be confusing, as they can be found naturally in text; it was important for the interaction indications to be distinct from the text.

For each of the 3 interfaces, everyone selected either red or dotted, or stated that they were about equal. Two of them mentioned that the red stood out more; one liked this, the other found it distracting. This choice changed somewhat from interface to interface, but people still listed red & dotted as the best choices. Plain (that is, no highlighting was done) was commented on as being very non-disruptive, but also not very helpful; no one selected it as a favorite for any interface.

In terms of preferred interfaces, 2 participants selected the ‘always highlighted’, and 1 selected the ‘toggled highlighting’. One participant who selected the ‘always’ interface said that the ‘toggled’ interface was decent, but that the button just added unnecessary complexity. The other ‘always’ participant disliked that the toggled interface disabled the highlighting, but not the tool-tips; he found it inconsistent that you could get tool-tips without any indication. The participant preferring the toggled interface mentioned that it was important to her to have the option; it’s possible that the highlighted text could be very distracting on some pages, and so it’d be essential to be able to disable it.

Overall, 2 of the 3 users wanted persistent indications that a word could be interacted with, even if it meant a little extra distraction. Yet all participants agreed that this indication should not change the shape of the text; very much in line with our previous observations on the importance of not changing the reading process. And of course we still see that people have different preferences; 1 liked the toggle button, others didn’t, 1 liked how red stood out, another didn’t. While the use of red or dotted text for highlighting was unanimously popular, the choice of the interface and the reasons why were not; once again we saw that different users have different perspectives.

7. Conclusions

In this project we set out to investigate the use of definitions embedded in text, to look at and evaluate several possible interfaces. Our initial survey indicated to us that there was substantial interest in such ideas, and through our online survey we were able to develop a general idea of what evaluation criteria were important, and what types of interfaces were seen as useful. Our first observational study provided a more in-depth understanding of these issues, and also allowed us to look at the issue of multilevel definitions. Finally we conducted an observational study in order to investigate different ways of highlighting interactive words, giving us information on a different facet of the issue.

There are 2 themes we’ve seen repeatedly in our studies. First, users tend to have substantially different preferences, that there is not necessarily one perfect solution, no single interface that would be ideal for every user. As we’ve seen in the online survey and both observational studies, a feature can be warmly received by one person and considered a nuisance by another. The design of an interface

must take this into account, perhaps via customizability or perhaps by targeting the least common denominator; in any case, developing an interface with only a certain type of user in mind is likely to alienate others.

Second, in order to be successful, new features should be added in such a way so that the fundamental reading process goes unchanged. Users appear to be willing to accept embedded definitions when they bring new functionality without any new demands, and when use of these features fits in naturally with the normal operation of the system.

8. Future Work

There are numerous ways we could extend this work; beyond larger scale studies, there are some additional aspects which would be interesting to explore. There are also a number of separate issues that would have to be addressed before an embedded definitions system could actually be deployed; we provide some initial thoughts on this in section 9.

8.1. Studies with Eye-tracking

In the work on Fluid Documents [6], the authors use an eye-tracker to better observe their subjects, and get a better indication of what they watch and for how long. Our observational studies allowed us to watch subjects as they read, and gather their comments, yet the act of commenting itself no doubt altered the process; through carefully designed tests based around an eye-tracker we may be able to gather data from more accurate instances of reading.

8.2. Other Areas

We could also consider the use of such a system in other areas, such as reading research papers, or, as one participant commented, in learning a foreign language. No doubt many of the ideas would carry over, but it is likely that other areas would require the system to be used in different ways, and so different interfaces may be called for.

9. Related Issues

This section will briefly address the outstanding issues that would need to be resolved before a full system could be designed. This project focused almost exclusively on interface issues, so a number of practical considerations remain. The big questions that remain include the source of definition information and the amount of configuration provided to the user.

9.1. Definition-Document Model

In our initial planning we considered schemes where the definitions are directly embedded in each document, as well as more modular approaches. There are arguments for each side, and we briefly mention the most evident ones below.

In our initial thinking, all definitions and associated meta-information would be directly incorporated in the file. This approach lends itself to offline usage because all resources are present in the file itself. This raises the likelihood that the provided definitions are helpful, and minimizes problems from multiple definitions because the author chooses the specific word. Unfortunately this requires an appreciable increase in effort for content authors with limited perceived benefits.

At the opposite end of the spectrum, users could download definition files that apply to a specific document, or more likely to a specific subject area. This kind of approach would be completely backward compatible, but the utility of definitions is more questionable.

A hybrid approach would include author annotation of words they felt needed definition, along with modular definition files. These systems would prefer author-specified definitions, but allow default definitions to be provided by alternate sites.

A final approach would be to automatically query publicly available sources, determine the best result and insert it on demand. We include this merely for completeness. At present, the natural language processing technology to support this appears to still be years away.

9.1.1. Origins of Definitions

The source of embedded definitions is likely to impact the usefulness of any system. In our observational study of multi-level definitions one participant complained about redundant information in a definition. Another said that unhelpful definitions would quickly decrease the perceived utility of the system. This agrees with our intuition that these changes will only enhance the user reading experience if they provide useful information.

In a deployed system, there would need to be a way to distinguish between definitions that are provided by the content author, those added by a third-party vendor, and amateur contributions. We did not explore the

implications of this issue or spend much time thinking about implementations. A confidence measure for definitions, possibly based on collaborative filtering, would be a good starting point.

9.1.2. Customizability

One of the main results of our studies was confirmation that people have different preferences about how definitions should be presented. In fact, several participants stated that the preferred interface would depend on the document in question. This helped to confirm our hypothesis that the ability to customize the definition presentation would be an essential part of any system. An entirely different project could be conducted just to design a suitable interface for customizing the definition presentation properly. Things to consider would be ease of activation/deactivation, range of options, and even the proper way to notify the user that definitions were available.

ACKNOWLEDGEMENTS

The authors would like to thank Kate Deibel for helpful comments, and all of the participants in our various studies.

REFERENCES

1. IntelliTXT. http://www.vibrantmedia.com/site/web 1a.html.

2. G. Marchionini and B. Schneiderman. Finding facts vs. browsing knowledge in hypertext systems. *IEEE Computer*, pages 70–80, 1988.

3. T. H. Nelson. Complex information processing: a file structure for the complex, the changing and the indeterminate. In *Proceedings of the 1965 20th national conference*, pages 84–100, New York, NY, USA, 1965. ACM Press.

4. Adobe Acrobat Reader. http://www.adobe.com/products/acrobat/readstep2.html.

5. P. T. Zellweger, B.-W. Chang, and J. D. Mackinlay. Fluid links for informed and incremental link transitions. In *Proceedings of Hypertext'98*, pages 50–57, June 1998.

6. P. T. Zellweger, S. H. Regli, J. D. Mackinlay, and B.-W. Chang. The impact of fluid documents on reading and browsing: An observational study. In *Proceedings of CHI'00*, pages 249–256, 2000.

7. Dictionary Search http://dictionarysearch.mozdev.org/