

Improved Readability and Information Access with Patent Documents: User Study of the Related Technologies

March 2, 2015

Abstract

abstract here

1 Introduction

Clear language is important in any human communication to ensure efficiency and eliminate risks of misunderstanding. With written text, this clarity, or ease-of-understanding, is measured by readability and it can refer to readability of a word, phrase, sentence, paragraph, section, document, or document collection.

In this study, we focus on readability of these text units in the context of patents. Their legal genre is difficult to understand in general (Alberts et al., 2011) and in particular to laypersons, that is, people without professional or specialised knowledge related to patents. Each patent document follows a predefined structure that consists of the title, abstract, background of the invention, description of the drawings, and claims, among other sections. For patent readability, the claims section can be seen both as the most important and most problematic section: they define the scope of legal protection of the invention and need to be written into a single sentence. This creates several readability problems: First, to protect the invention, wordings are very carefully considered, resulting in difficult-to-understand legal jargon - sometimes referred to as “lawyers French” - both in terms of vocabulary and grammar. For example, *consist of* in a patent claim is defined to be a *closed transition* and before a list, it means that inventions infringing this claim need to have all and only the listed elements/steps. In contrast, *consisting* is a *hybrid transition* with the specific meaning of other inventors possibly being able to invent something new (without infringing this claim) by employing some additional elements/steps. Second, the sentences are so long and grammatically

hard, that understanding their content becomes difficult. The third transition type is called open transition, defined as having at least the following elements/steps. To illustrate, let us provide the following 156 words example claim sentence with the three types of transition phrases underlined:

Toolholder, which could comprise of a holder body with an insert site at its forward end consisting essentially of a bottom surface and at least one side wall where there projects a pin from said bottom surface upon which there is located an insert having a central bore, clamping wedge for wedging engagement between a support surface of the holder and an adjacent edge surface of said insert and an actuating screw received in said wedge whilst threadably engaged in a bore of said holder, said support surface and said edge surface are at least partially converging downwards said wedge clamp having distantly provided protrusions for abutment against the top face and the edge surface of said insert, characterized in that the wedge consists of a pair of distantly provided first protrusions for abutment against a top face of the insert, and a pair of distantly provided second protrusions for abutment against an adjacent edge surface.

Similarly to our example of words with specific meaning, both punctuation and lexical markers carry a specific meaning (to a patent lawyer but not to a layperson). Furthermore, a claim should be composed by, at least, the following parts,

- **Preamble:** a preamble is an introduction, which describes the class of the invention.
- **Transitional phrase:** a transitional phrase is a phrase or linking word that relates the preamble with the rest of the claim. The expressions comprising, containing, including, consisting of, wherein and characterise in that are the most common transitions.
- **Body:** the body text describes the invention and recites its limitations.

Our goal is to study laypersons' attitudes towards technologies for improving patent readability. By following the definition of the Oxford dictionaries (<http://www.oxforddictionaries.com/>) we use laypersons as a reference to those without professional or specialized knowledge in patent documents.

Our research method is a questionnaire study, implemented as a cascaded approach of using different technologies and answering our online questionnaire. This cascade is divided to the following five readability areas with six tested problems in total:

1. **Word readability:** Understanding a patent claim because it has jargon words with special meaning. (Test 2)
2. **Claim readability:** Understanding key topics in a patent claim because it is an unusual long sentence. (Test 3) Understanding patent claim because they are unusually long. (Test 1)
3. **Section readability:** Understanding claim dependency structure. (Test 4)
4. **Document readability:** Understanding patent classification codes that summarise the document content. (Test 5)
5. **Collection readability:** Retrieving a collection of relevant patent documents from the gamut of documents, understanding this collection, navigating through it, and summarising it in terms of (Test 6) its topics and topic relations and its authors and organisation and their relations.

See <https://docs.google.com/document/d/1SoaVrwgoImt25hrgrNM9DZ5ihJj7aGunm/edit#forfurtherelaborationoftheseproblemsandsolutionontechnologies>.

2 Materials and Methods

2.1 Theoretical Background

The theoretical background of this study is based on the papers related to the technology acceptance model (TAM) (Davis et al. 1989; Venkatesh and Davis 2000), task technology fit (TTF) (Goodhue 1995), and web site user satisfaction (WSUS) (Muylle et al. 2004).

TAM (Davis et al. 1989) attempts to understand why people accept or reject information technologies. It applies a very general model of human behaviour called the theory of reasoned action (TRA) (Fishbein and Ajzen 1975) to the domain of user acceptance in the context of information technologies. This serves as a theoretical justification for specifying causal linkages between perceived usefulness (PU), perceived ease of use (PEU), and the concepts of the Structural Equation Model (Igbaria 1997): users attitudes, intention, and technology adoption behaviour. TAM defines PU is defined as the degree to which a person believes that using a particular technology will enhance his/her work performance and PEU as the degree to which a person believes that using a particular system will be free of effort. TRA postulates that peoples reasoning flows from 1) their beliefs and evaluations through 2) them developing an attitude towards performing a certain behaviour via 3) this attitude then having a causal role with their intention to perform this behaviour to 4) this intention finally defining if the behaviour is executed

or not. In this model, external variables (e.g., user characteristics, political influences, organisational factors, and development process) are expected to influence technology acceptance indirectly by affecting peoples beliefs, attitudes, or intentions (Szajna 1996).

We chose this model, because TAM and/or its key concepts of PU and PEU are widely accepted among the research community as tools for evaluating technologies and predicting their use and are applicable to cases such as ours where technology use is optional (Doll et al 1998). TAM has also been validated with a number of technologies (Davis et al. 1989, 1992; Davis 1993; Igarria 1993; Igarria et al. 1994; Dishaw and Strong 1999; Horton et al 2001) and cultures (Straub et al. 1997). Moreover, TAM can be used to evaluate technologies very early on in their development or assess user reactions to technologies on a trial basis (Davis et al. 1989). Finally, only a very brief period of user interaction with technology is needed before the model is capable to explain and predict user acceptance (Szajna 1996; Doll et al. 1998).

TTF (Goodhue 1995) views technologies as means for a goal-directed person to perform tasks. It posits that technologies will be used if, and only if, their available functionalities support the users activities. Consequently, its focus is on the match between user task needs and available functionalities of a given technology. The match is measured by the extent that 1. technology functionalities or characteristics meet its 2. users abilities and 3. task requirements (Goodhue 1995, Dishaw and Strong 1999). In other words, the model is three dimensional. Rational, experienced users will choose to use technologies that enable them to complete their tasks with the greatest net benefit and not to use technologies that will not offer them a sufficient advantage. Similarly to TAM, TTF is applicable to cases where technology use is optional. TTF has been also integrated with TAM before, because these two models have a clear overlap (Dishaw and Strong 1999). This integration considers both the attitudes towards technologies and the fit between technology functionalities and task characteristics relevant to a given user. According to the integrated model, the three parallel dimensions affecting the fourth dimension of 4. user evaluations of TTF are 1. technology characteristics, 2. users skills and abilities, and 3. interaction between the task and technology (and user). This has been found to offer a significant improvement over either TAM or TTF alone.

We chose TTF because our goal is to assess which technologies/technology functionalities contribute to the readability of patent claims, documents, and document collections from laypersons perspective in the context of reading and searching tasks. We decided not to choose the integrated model but rather to consider both TAM and TTF separately, because this seems to be more common practice in the research community.

WSUS (Muylle et al. 2004) is based on a two-step study. First, a pilot study

was conducted in order to define which items contribute to user satisfaction if using a web interface to interact with technology users. The contributing items were along eleven dimensions (i.e., 1. information relevancy, 2. information accuracy, 3. information comprehensibility, 4. information comprehensiveness, 5. ease of use, 6. entry guidance, 7. website structure, 8. hyperlink connotation, 9. website speed, 10. layout, and 11. language customisation). The dimensions 1-4 constitute the component of information, the dimensions 5-9 the component of connection, and the remaining components are layout (i.e., dimension 10) and language customisation (i.e., dimension 11). Second, a confirmatory factor analysis was performed for these items and its results demonstrated the adequate validity and reliability of the initial model. The theories laying the foundation for WSUS relate to measuring technology users computing satisfaction (Doll and Torkzadeh 1988) with five measurable dimensions that correlate with user satisfaction (i.e., 1. content, 2. accuracy, 3. format, 4. ease of use, and 5. timelines); information success model (DeLone and McLean 1992) with six dimensions (i.e., 1. system quality for measuring technical success and 2. information quality for measuring semantic success together with 3. use, 4. user satisfaction, 5. individual impacts, and 6. organizational impacts for measuring effectiveness success); and slightly revised information success model that takes changes in the role and management of technologies in 1990s and early 2000 (DeLone and McLean 2003).

We chose to consider WSUS, because patent search is typically implemented on the Internet as a web search engine (e.g., The Lens (<http://www.lens.org/lens/>) a open public resource of this kind). It is also a contrasting approach to the supplementary models of TAM and TTF.

2.2 Questionnaire

Our questionnaire consisted of XXX questions in English (Table XXX). XXX of the questions were unstructured so that the participants could freely write down their own comments. Other questions were structured and the answers used a five- or three-point scale analogous to the Likert scale.

The questionnaire was implemented using the open source on-line survey application: LimeSurvey (<https://www.limesurvey.org/en/>)

Questions were designed using the papers by Davis (1989), Goodhue (1995), Doll et al. (1998), Dishaw and Strong (1999), Venkatesh and Davis (2000), Anandran et al. (2000), and Muylle et al. (2004). To incorporate elements of interaction and visual analytics that are crucial in the Internet technologies to support retrieving a collection of relevant patent documents from the gamut of documents, navigating through it, summarising it, and understanding its each claim in 2010s, we included some timely techniques from user evaluations related to search en-

gines (Angelini et al. 2013; Hill and Toms 2013). However, all questions proposed in the papers were modified so that they fit the tasks at hand. To shorten the answering time, the questions were also formulated in a way that the same question fits at least two of the theoretical models. All questions and the related models are presented in Table XXX. The question pattern is repeated separately after each part of the six-step study (we refer to the parts with the term Test (See Section 1)).

The original studies were used to formulate our questions. TAM questions were based on Davis (1989), Doll et al. (1998), Dishaw and Strong (1999), Venkatesh and Davis (2000), and Anandraj et al. (2000). However, some questions were combined to keep the number of questions as small as possible, but still paying attention to covering all dimensions of the original model. TTF questions were based on Goodhue (1995) and Dishaw and Strong (1999). WSUS questions based on Muylle et al. (2004) did not address the dimension 6. entry guidance at all, because our test modules do not have a joint entry page yet and also our technologies address English laypeople reading English patent documents only. Moreover, they did not address the dimension 9. website speed either, because our technology development has rather aimed to optimise the content than speed; after validating the content through this user study, we will address the speed dimension. Finally, in order to shorten the answering time, we chose only approximately one question for each WSUS dimension as opposed to the models' original three to four questions per dimension.

Gabi said: add column for answers in Table 1 ?

2.2.1 Demographic questions

1. What is your age?

- 20 or younger
- 21-30
- 31-40
- 41-50
- 51-60
- 61 or older

2. What is the highest level of school you have completed or the highest degree you have received?

- Less than high school degree
- High school degree or equivalent

Table 1: Questions and related models

Questions and answer options	TAM	WSUS	TTF
In my opinion, it is important to improve the readability of patent documents: strongly agree / somewhat agree / neutral / somewhat disagree / strongly disagree			
In my opinion, it is difficult to read patent documents: strongly agree / somewhat agree / neutral / somewhat disagree / strongly disagree			
Please order the ways of representing the patent document of Test t in your preference order: 1. my most preferred option is: N. my least preferred option is: ...			
In my opinion, there is an even better way of representing the patent document: yes / no . If yes, please describe it as free text.			
I do not find information I need easily from patent documents as they are now: yes / no			
Please, explain what aspects of the way patent documents are presented are you happy with.			
Please, explain how this format could be improved			
I think my preferred option supports reading patent documents			
I can see myself using my preferred option when reading patent documents			
I plan to use my preferred option (given it was made available			
I think my preferred option supports my work			
In my opinion, I do not find information I need easily from patent documents even if using the most preferred option			

- Some college but not degree
- Bachelor degree
- Graduate degree

3. Which of the following best describes your current occupation?

- Business and financial operations occupation
- Computer and mathematical occupation
- Architecture and engineering occupation
- Life, physical science
- Community and social service occupation
- Legal occupation
- Education, training occupation
- Arts, design, entertainment occupation
- Health care occupation
- Office and administrative occupation
- Other (please specify).....

4. Have you ever read a patent document?

- Yes
- No

5. Have you ever write a patent document or part of it?

- Yes
- No

6. Do you use or have you used patent documents at your work on a daily basis? If yes, how and when?

- Yes
- No

2.3 Tests used

Gabi said: include figures with the tests from the survey Jaume?

Table 2: Participant demographics

Question	Percentage
What is your age?	
20 or younger 0	0.00%
21-30	13.11%
31-40	36.07%
41-50	22.95%
51-60	14.75%
61 or older	4.92%
No answer	8.20%
What is the highest level of school you have completed or the highest degree you have received?	
Less than high school degree	0.00%
High school degree or equivalent	4.92%
Some college but not degree	8.20%
Bachelor degree	22.95%
Graduate degree	65.57%
Which of the following best describes your current occupation?	
Business and financial operations occupation	9.84%
Computer and mathematical occupation	34.43%
Architecture and engineering occupation	4.92%
Life, physical science	3.28%
Community and social service occupation	3.28%
Legal occupation	8.20%
Education, training occupation	14.75%
Arts, design, entertainment occupation	14.75%
Health care occupation	4.92%
Office and administrative occupation	6.56%
Other	9.84%

3 Evaluation results and discussion

3.1 Participants

[Gabi said: should we explain here the recruitment process?](#) Recruitment: XXX

We got 65 participants in total, of which XXX were by men and XXX by women (see Table 2 for more detailed information about the population). There were XXX (education and affiliation descriptives). However, because we did not build any statistical models using these descriptive categories as, for example, grouping variables, we decided not to detail the analysis on this level either.

[Gabi said: note that we did not ask for the gender](#)

Some (7 out of 65) participants gave more details about their occupation, and the categories are: patent examiners, journalist, translator, cultural and information

Table 3

Question	Count	Percentage
Have you ever read a patent document?		
Yes	35	57.38%
No	19	31.15%
No answer	7	11.48%
Have you ever written a patent document or part of it?		
Yes	10	16.39%
No	45	73.77%
No answer 6	9.84%	
Do you use or have you used patent documents at your work on a daily basis? If yes, how and when?		
Answer	35	67.31%
No answer	17	32.69%

occupation and management.

3.2 Analysis of the Answers

All analyses were performed by using the statistical software package XXX.

Eight out of sixty five said they used patent documents at work on a daily basis.

4 Conclusion and future work

Gabi said: [Hanna, do you have the references in bibtex?](#)