Qualitative approach to understanding context when people compare and integrate two websites ?

Authors

Arindra Kumar Das <akdas@kth.se>

Weiwei Zhang <weiweizh@kth.se>

Arindra Kumar Das is responsible for literature review, setting up the experiment, carrying out the interview, analyzing the data and presenting the project.

Weiwei Zhang is responsible for investigating background and literature review, data collection during observation and interview, data analysis and presenting the project.

Abstract

Today's websites have information ambiguities and with the Semantic Web we can reduce the severity of this issue. The Web Ontology language (OWL)[11] deals with content mapping of such websites, but heterogeneities make it insufficient[14]. Keeping this problem in mind, we strive to gain insights of how people contrast and integrate two websites in order to find possible improvements for information integrating algorithms. Humans behave and perceive knowledge in various ways. To model a machine readable language according to human thinking is quite challenging. By performing an experiment and analysing data, we finally form a step-wise analytical model which represents people's strategy and context when compare and merge two websites. There are seven kinds of context discovered i.e. background knowledge, previous experience, first glance, culture, personal preference, pattern and meaning.

Keywords: Context, Web Ontology Language (OWL) algorithm, Website merging, Semantic Web

1. Introduction

The World Wide Web is a repository for a lot of information. Enterprises, government and many other organizations, etc. provide information and services to promote their operation via websites. However, this information is not managed in a proper way, as it overlaps, not to mention the inconsistency that exists between web pages owned by the same organization. Using the Semantic Web, which forms a meaningful content and structure for machines to retrieve information online [1], this problem can be solved efficiently.

1.1. Background

The Web Ontology Language (OWL) [11] is a language that aids conceptualizing information online and gives a better representation of websites based on Semantic Web. Problems still exist in different ontology matching systems because of heterogeneities [14]. According to the English dictionary, the contextual information is a combination of linguistic and non-linguistic context. In our project, linguistic context of a website refers to the contents and the surrounding white spaces around them, while non-linguistic context includes the designs, patterns, layout, color, styles etc. of a website, and also includes people's background knowledge and their understanding of information they perceived.

The human mind processes information by understanding the events and patterns around it [12]. Finding out what factors people consider when comparing and merging two websites can help develop machine readable algorithms to simulate human behavior.

1.2 Problem

We seek to discover how people think and act, when mapping similar content of two websites and try to find helpful information to enhance the algorithm behind OWL.

In this project we will investigate how people perceive and understand information on websites with linguistic and non-linguistic context.

Our problem is: *What can be used as context when people compare and integrate two websites.*

1.3 Purpose

In this report, we present readers with our findings from an experiment in which we find certain patterns and relevant context that people use when comparing and merging two websites. By explaining and validating our research method, data collection, and analysis method, we hope that the readers will find the outcome of our research useful for improving OWL algorithm.

1.4 Goal(s)

Our goal is to derive an analytical model that represents the process of people's perception of semantics and patterns in various websites and the strategy they use so as to compare and merge them.

2. Method

In this project, we used *qualitative research method* because the problem which we look at concerns human psychology and behavior. Unlike *quantitative research method*, we collected data like perceptions, users' understanding and opinions instead of numbers [3].

We followed the *interpretivism assumption*. As we investigated how people assign meanings to things (websites), how they see and what context was involved in the whole procedure, thus the object of our study involves individual's opinions and perspectives.

Further, we chose *empirical research method*, because we aimed to summarize people's thoughts and opinions, which are affected by one's experiences.

Our project is experimental and we did not test any theories, instead, we wanted to find all the factors that affect participants in the process of merging. Therefore, we followed *inductive research approach* and *exploratory research strategies*.

In order to collect data as opinions and experiences, we chose *Participant observation and semi-structured interviews* for data collection rather than *questionnaire or survey.* It was because through close observations and interviews we gained insights of participant's opinions, which could not be gained from questionnaire or survey.

Afterwards, we analysed the data using *narrative analysis method*. However, we struggled initially between *analytic induction* and *narrative analysis.* This was because we were interested in certain contexts when people compare two websites. And *analytic induction* deals with the context of specific patterns. But, *analytic induction* seeks to find a universal causality between specific context and phenomena through iterative case study, which did not fit our case. Since we tried to make the subject open up, we agreed that every person is unique and we did not know if a universal cause-and-effect relationship existed. Also, we were curious about how would people interpret the similarities between websites in their own ways. Thus, we thought of *narrative analysis*, through which we could study what and how people say and act. More specifically, we chose the *labovian analysis* method since it is used to form a structured sequence of causal events, from which we could clearly see the function of each sentence in participant's narratives, thus easily found out the context they used during the experiment [4].

In narrative analysis, the researcher is required to establish and an intimate relationship with participants, as the quality of the information partly depends on the relationship between researchers and participants [8]. Due to limited time, we invited four of our friends as participants. We believed our closeness enables a higher possibility that they would express their opinions truly without stress and we would interpret their behaviours, sentences more correctly. The participants are of different genders and from two different knowledge areas to see if these factors would affect how they accomplish the task.

Before the experiment, we explained our participants what that entailed and why we did this project. We assured them that the data collected in this experiment would be used only for the research purpose of this project, and would not include their private information [13]. We started the experiment only after we got consent from the participant *(see Appendix f)*.

Since we collected our data through observations and interviews, the replication of this data is hard, as the interpretation of observation will vary with different individuals. Also, it is hard to control the content of the interview because each interviewer might give different response to the same question. It is even harder to control it in a semi-structured interview, as more questions arise in the process based on the interaction between the interviewee and the interviewer.

In the case that the interview is directed to irrelevant areas by the interviewee, we needed to focus on our problem: the method by which they fulfill the merging task. We used imperceptibly control to lead the participant back to the topic when it is necessary[2].

As a result, our analytical model, is the outcome of a sequence of data collection and data analysis under existing rules, thus proving the validity of our research [3].

2.1 Data collection

Our participants' task was to compare and merge a pair of similar websites we chose. Participants were free to use any kind of help and merge the websites in any form they wanted.

We chose *The Times* and *The Sunday Times* as the websites for our task because they were both news websites managed by the same company with different focus on information. Both websites had similar as well as different information. This provided a good opportunity for us to see how the participants deal with information of various kinds and relationships. Additionally, they had similar structures, which made it easier to compare.

Considering that both websites have a huge amount of information, we asked participants to merge a certain part of the websites to avoid the effect of negative emotion if he feels overwhelmed and tired.

Our experiment was held at EIT ICT Stockholm Co-location Center, a comfortable place for students to study, in order to give our participants a familiar and casual environment.

When our participant was doing the merging task, we actively observed the participant and asked questions when they were willing to tell [5]. Meanwhile, we kept field notes separately under the observation guidance

introduced by Merriam [10]. This process was recorded by two cameras from different angles for later confirmation with our observation notes.

When the participant finished the task, we conducted a semi-structured interview with him with three main questions (*see appendix e*) followed by open questions. The interview was also recorded as an aid for the pre-construction of participant's narrative which was done after the interview.

2.2 Data analysis

During the observation, we kept field notes [9] of the settings, participants' relevant activities and our instantaneous understanding. These data were then thoroughly analysed after validating the field notes with the video and audio recordings.

In the first two experiments, we observed participants with a focus on subtle emotions, detailed movements and talks. While in the last two experiments, we improved our observations by focusing on how they actually fulfilled the merging task. In other words, we focused on participants' strategies and how various factors were linked and affected their decisions. Our post-observation analysis is shown in appendix c *(Table 1, 2, 3 and 4).*

For the data collected from the interviews, we did pre-construction [6] before following the process of labovian analysis, which has become a standard narrative analysis method [4]. Then we analysed upon the massive amount of structured data *(see Appendix b)*, which were the results of labovian analysis. We removed irrelevant information and focused on participant's strategies, what context they used and the evidence from their narrative which we believed can prove why they use these contexts.

3. Results

From the analysis, we found seven different contexts which participants used during the whole experiment: background knowledge, previous experience, first glance, culture, personal preference, pattern and meaning. Our post-interview analyses are shown in Appendix d *(Table 5, 6, 7 and 8).*

Our analytical model *(shown in Figure 1, 2, 3, 4 and 5)* represents the process of how our participants compare and merge two websites. It comprises a sequence of 5-steps (*Step a, b, c, d and e*). Each step includes evidence from participant's data analysis and improvement in OWL algorithm based on semantic Web. The evidence is presented with the related contexts which we interpreted from our analysis.

Step a of the analytical model presents that the participants started by getting information from two websites. Participants 1 and 3 used personal preference as the context to find underlying structure of the website, purpose and functions of each part. Previous experience and culture were the contexts used to get the most important information from the website by participant 2, who explained that in her country people read from left to right. So she tended to look for most important information at the middle-left part of the website. Participant 4 started by merging and designing her own website.

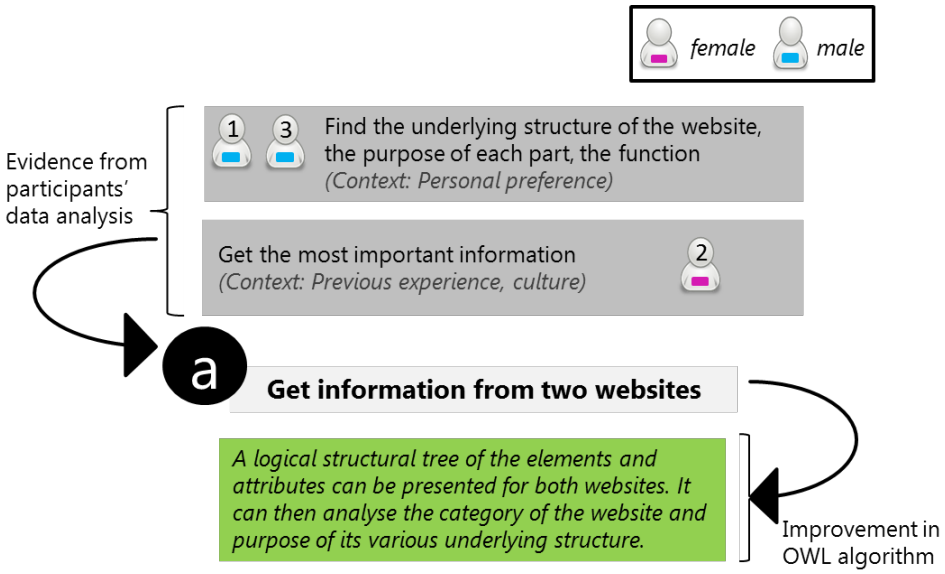

Figure 1. Step a: First step of the analytical model discusses about how the information is obtained from two websites.

Step b of the analytical model represents how the participants tried to find similarities between two websites. Previous experience and culture were used as contexts by participants 1 and 3 to find similar purpose of each part. (e.g. They explained that they used to read online news a lot, so they knew what was the purpose of a certain part of the website with some specific titles.） By using previous experience and background knowledge,

all participants found similar structure of two websites.

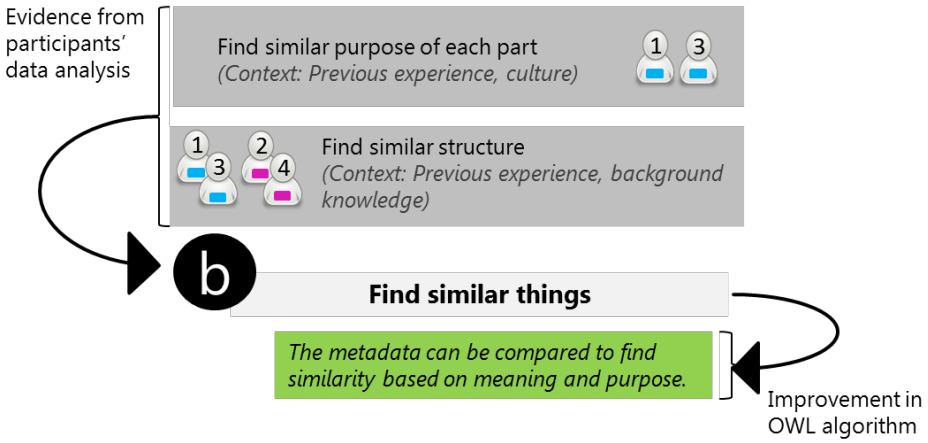

Figure 2. Step b: Second step of the analytical model discusses about finding similar things in the two websites.

Step c shows how participants decide what to do with the similar part. So, based on similarity we divided the evidence from participants' data analysis into two categories. If they are of the same content and structure, participants simply merged them into one and if they are similar but not the same, there were three different situations to be discussed. First, if participants thought of them as parallel using contexts such as meanings and previous experience (e.g. they have seen similar categorization before in other websites), they were put into a bigger group. Participants used previous experience, culture and background knowledge to decide if they had child- parent relationship in order to merge one under another category. Thirdly, in situations where participants concluded by using previous experience that the similar parts were related but cannot be put into one category, participants kept them together.(e.g. participant 4 made

a category called review & opinion in new website).

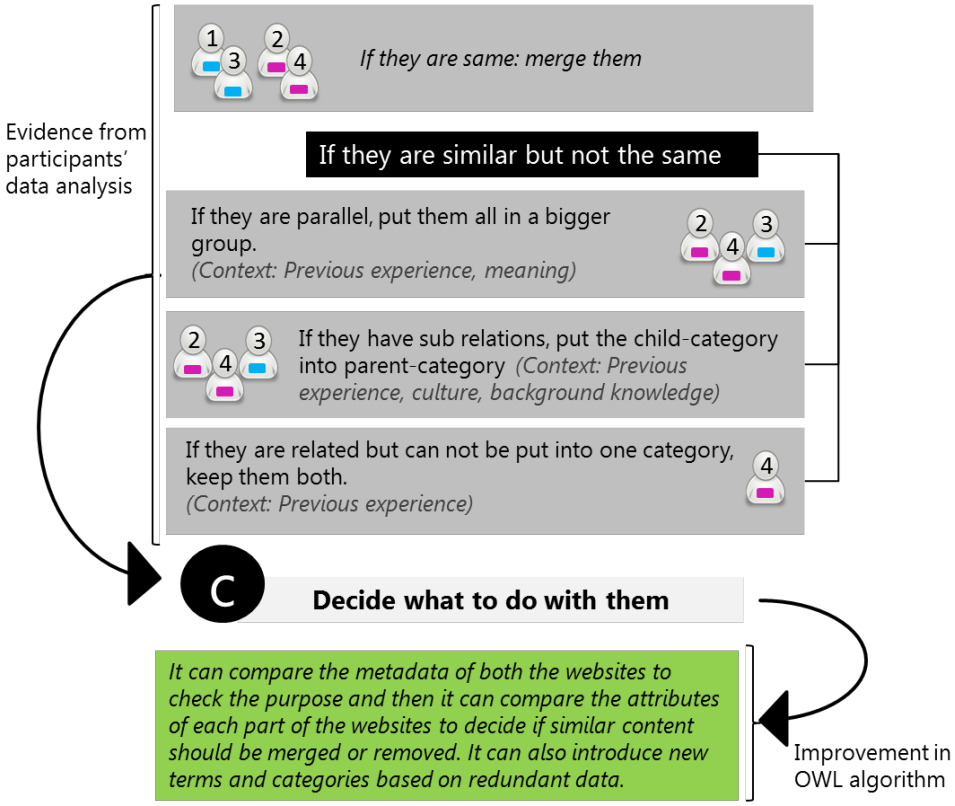

Figure 3. Step c: Third step of the analytical model discusses about deciding what to do with similar things in the two websites

Step d of the analytical model shows how participants deal with the rest of the websites. Participants used previous experience as context to find out the importance of each part of the website and kept them accordingly in the merged website. (e.g. participant 4 said that opinions and reviews are important in news websites, as she had seen it in most newspapers, so that she wanted to keep this part at a upper position). Also, all participants used pattern, meaning and previous experience as contexts to figure out the relevance of each part to the purpose of the websites and assigned spaces to them according to their relevance.

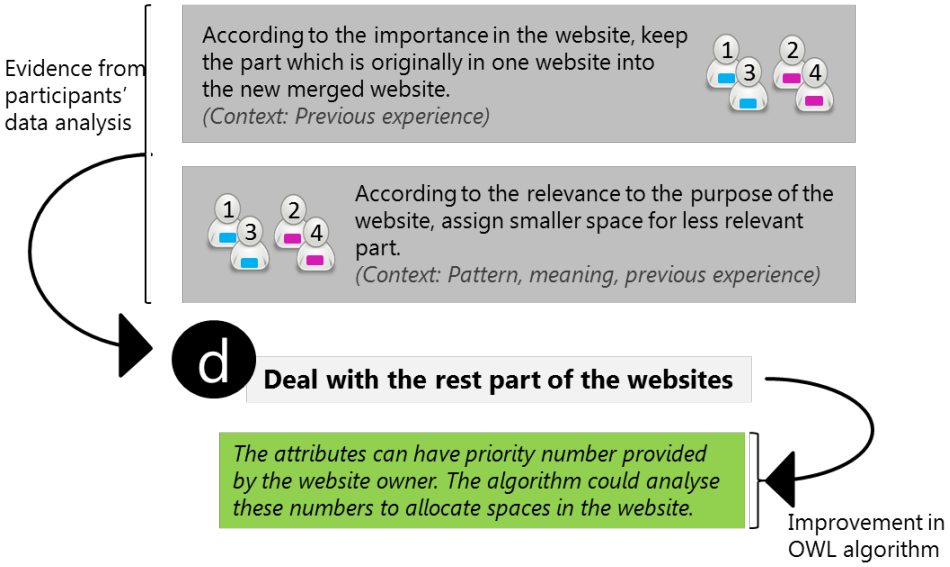

Figure 4. Step d: Fourth step of the analytical model discusses about dealing with the rest part of the websites.

Finally in Step e, participants chose the style, design of the new website using contexts as personal preference and background knowledge (e.g. participant 4 wanted to design a totally new website as she explained what she thought news websites should be like, which was based on her experience of reading newspapers. Participant 2 and 3 preferred the style of one website more because it tended to have more design value)

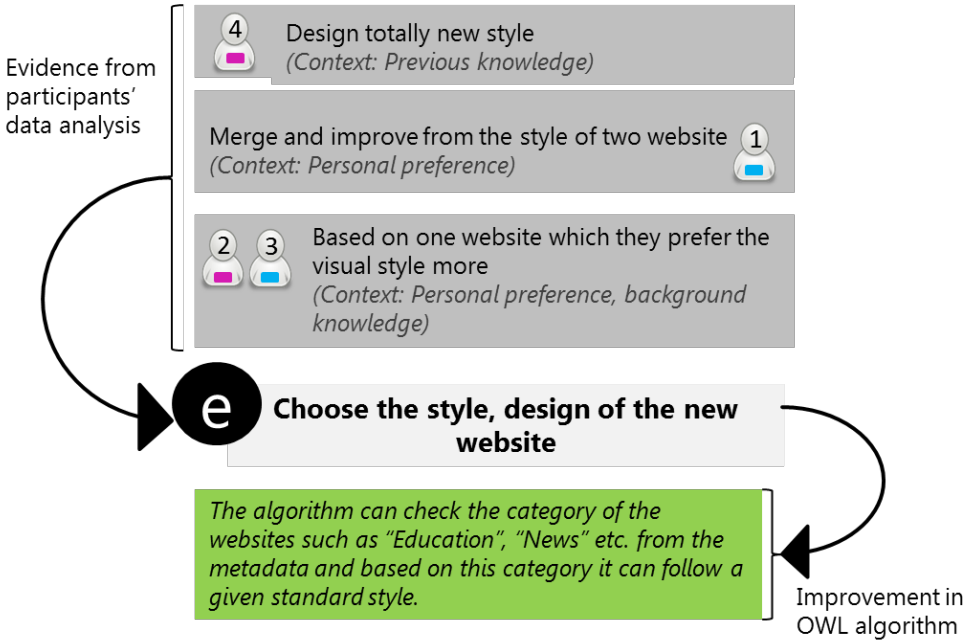

Figure 5. Step e: Fifth step of the analytical model discusses about choosing the style, design if the new website.

3.1 Discussion

The result derived from the four participants reflects many different contexts based on individuals, such as background knowledge, previous experience, first glance, culture, personal preference, pattern and meaning.

In our research, the background knowledge is the knowledge within a certain academic area of participants' study (e.g. Usability and Design); Personal preference is the participants' formed opinions and tastes; First glance is the participants first visual impression of the website; Culture is the participants societal convention reflecting their way of doing things; Previous experience is participants' experiences that are stored in their memory, which they use as reference to perform similar tasks; Pattern is a regular way to arrange similar contents in a news website which was unconsciously learned by the participant. Finally, we explain the the context 'meaning' as follow: when the participants compared the elements of two websites, they tried to understand all the necessary information, which includes purpose, function (see *Table 7 in Appendix c)*, and their understanding of certain words *(see Table 8*).

Contexts like personal preference and culture cannot be directly interpreted into the OWL algorithm. However, in Semantic Web, we can translate the contexts into machine readable information. [7]

4 Conclusions and Future Work

In this paper we investigated the process of how people comparing and merging two websites by doing experiments with four participants with two genders and two knowledge background. We found seven contexts they used during this process, i.e. background knowledge, previous experience, first glance, culture, personal preference, pattern and meaning. And we presented the result in an analytical model.

As we followed certain rules in every phase of our research, it is possible for others to do similar experiments, if they abide by the same rules [9]. If other researchers find the same contexts, the correctness of our project will be approved, whereas if they find other contexts, our model will be expanded.

So far, our analytical model is just a starting point, if more experiments are implemented, it might be more useful for improving OWL algorithms.

Word count: 2559

References

[1] Berners-Lee, T. ; Hendler, J. ; Lassila, O.(2001) *"The Semantic Web"*Is Part Of: Scientific American, May 2001, Vol.284(5), pp.34-43

[2] DiCicco-Bloom, B. , Crabtree, B.F., (2006) *"The qualitative research interview"* Blackwell Publishing Ltd 2006. MEDICAL EDUCATION 2006; 40: 317

[3] Håkansson,A., "*Portal of Research Methods and Methodologies for Research Projects and Degree Projects*." WORLDCOMP'13 - The 2013 World Congress in Computer Science, Computer Engineering, and Applied Computing, 22-25 July, 2013 Las Vegas, Nevada; USA. (this paper will be in Bilda)

[4] Hunt J., (2006) *"Note of Labovian Analysis"*, Available at: http://www.eventyr.co.uk/2006/10/labov.html , Accessed September 25, 2013.

[5] Kawulich, B. B. (2005). "*Participant Observation as a Data Collection Method"* [81 paragraphs]. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 6(2), Art. 43, http://nbn-resolving.de/urn:nbn:de:0114-fqs0502430, Accessed October 21, 2013.

[6] Labov, W., *"Narrative pre-construction"* Available at http://www.ling.upenn.edu/~wlabov/Papers/NPC.pdf, Accessed October 21,2013.

[7] Lamas, D., (2011), *"Introduction and theoretical foundations of New Media: Metadata and Ontologies"*, Available at: http://www.slideshare.net/davidlamas/metadata-and-ontologies, Accessed October 21, 2013.

[8] Lisa M. Given (2008) *"The SAGE Encyclopedia of Qualitative Research Methods"*

[9] Marshall, C. & Rossman, G. B. (2006) *"Designing Qualitative Research Fourth Edition"*, p. 101

[10] Merriam, S. B. (1998). *"Qualitative research and case study applications in education"*. San Francisco: Jossey-Bass Publishers.

[11] Motik, B., Grau, B.C., Horrocks, I., Wu, Z., Fokoue, A., Lutz, C. *"OWL 2 Web Ontology Lang"*. W3C, August 2009, Available at http://www.w3.org/2007/OWL/draft/ED-owl2-profiles-20090420/all.pdf , Accessed September 9,2013.

[12] Norman, D. A. *"The Design of Everyday Things 1st edition"*. New York: Basic Books, 2002, ISBN: 9780465067107

[13] Orb, A., Eisenhauer, L. ,Wynaden, D., (2000) *"Ethics in Qualitative Research"* JOURNAL OF NURSING SCHOLARSHIP, 2000; 33:1, 93-96. ©2001 SIGMA THETA TAU INTERNATIONAL.

[14] Wu, D. *"Ontology Integration with Non-Violation Check and Context Extraction"*, Licentiate Thesis, 2013, Available at http://kth.diva-portal.org/smash/get/diva2:602098/FULLTEXT01.pdf Accessed October 10,2013.

Appendix

Appendix a. Milestone chart

The project will start on 14 September and end at 00:00 on 21 October. There will be the following milestones and deliverables:

Task	Start Date	End Date
Literature Study about our topic	14 September 2013	20 September 2013
Researching on method description	20 September 2013	02 October 2013
Background study on selected methodology	02 October 2013	06 October 2013
Experimental preparation	06 October 2013	08 October 2013
Data collection and pre-construction	08 October 2013	16 October 2013
Data analysis and comparing	16 October 2013	17 October 2013
Analytical method	17 October 2013	19 October 2013
Conclusion and future work	19 October 2013	20 October 2013
Refining the report	20 October 2013	21 October 2013
Opposition	21 October 2013	23 October 2013
Final presentation	23 October 2013	25 October 2013
Final report	25 October 2013	06 November 2013

Appendix b. Preconstruction and Labovian analysis

Due to a massive amount of our data analysis, we only provide an example to illustrate our methodology:

(1)

Pre-construction

"lifestyle" "be coz your lifestyle depends upon your culture and your style defines your lifestyle and according to travel your lifestyle will change" "so it depends on all the three things"

From the evidence of narrative as participant produced on asking the question about "*What word do you use to merge, travel,culture and style ?*" in (1), the chain of events he remembered would appear as (2).

(2)

Labovian analysis

e-$_0$:She thought that lifestyle depends on all three things

e-$_1$:She thought by travelling one's lifestyle will change, one's style defined one's lifestyle and one's lifestyle depends upon one's culture.

e-$_2$: She thought lifestyle is the proper word to merge travel, culture and style.

Abstract: "lifestyle"

Evaluation: "becoz your lifestyle depends upon your culture and your style defines your lifestyle and according to travel your lifestyle will change"

Coda: "so it depends on all the three things"

Appendix c. Post-observation analysis

Observation	Interpretation
He talks about “Malala Yousafzai :a news topic in the website” three times during the experiment and he also commented about few sports news	The participant considered both the content and the layout of the website when he compared them, a higher proportion is focused on the content.
He wants to delete that part “Subcategories” “Quicklink” when merging them because he didn’t understand it and he saw it as nonsense.	The participant was sensitive with the meaning of the words and the category.
He recalls when he was young in India, he used to fight with his siblings for the Sunday Newspaper, because according to him it used to have many cool stuff.	The participant perceived the information on the websites in a similar way as he used to read newspaper.
He comments about the Navigation, width of the website and home button	The participant cared about how information is presented and organized because he had seen a lot of different websites.
He concludes that The Sunday Times is newly developed because it use new technology in slide show of news.	He sees technological information

Table 1: The data analysis of the observation from Participant 1(male) from a communication engineering background.

Observation	Interpretation
She took notes, draws wireframe and scroll very slowly to view and analyse the two websites	The participant is very conscious and view the two websites in detail.
She complains about the color scheme that looks similar between two different sections which makes it hard to differentiate, she talks about the meanings behind the colors: pink for girl, green in sport means a football team in Greece, blue means sports to her.	The Participant cares more about structure and color and didn't mention anything about contents of the two websites.
She used terminologies in HCID "Consistency", "Feedback", "Constraint" while comparing the two websites, she noticed if there is color matching in the navigation and the heading of the news, and used Marketing terminologies when merging them .	The participant used her academic knowledge when she compare two website. she recently watch online course about entrepreneurship so she used marketing terminologies.
She said the navigation in The Sunday Times is easier to see, but she still likes the navigation in The Times because it's more beautiful	The participant cares about the design and feel of the websites
She concludes that The Sunday Times is newly developed because it have more images and large images)	The participant saw pictures as information

Table 2: The data analysis of the observation from Participant 2 (female) from an usability and design background.

Strategy	Context	Context

	Preference(evidence)	**Previous Experience (evidence)**
Remove the items		
• Home icon		‘This is a news website, I will keep the news” (not home icon)
Keep items		
• Keep the design of The Times	He likes the visual design of The Times.	
• Keep ‘quick list’ and ‘latest news’ in merged website	He think they are different but both very important	
• Keep ‘car/find/job’		
Categorise, compare, merge and change		
• merge ‘comment’ under ‘opinion’	Comment is a form of opinion	“opinion is more often used”
• merge ‘culture’,‘style’,and ‘travel’ into a new category ‘life’		“culture is part of life, and style, travel can also be under life”
• change ‘puzzle’ into ‘game’		chess and bridge is not really puzzle
• merge ‘magazine’ and ‘paper’		
• merge ‘driving’ under ‘sports’		
• combine ‘twitter’ and ‘blog’	The are all about social networks	

• combine some subcategories	They are less important.	

Table 3: The data analysis of the observation from Participant 3 (male) from an usability and design background.

Strategy	**Context**	**Context**
	Preference (evidence)	**Previous Experience (evidence)**
Make her own navigation	didn't like the navigation of both site	
Remove the items		
• list guides	She didn't like the list guides in The Sunday Times	
Keep the Heading in the middle	She like the heading in The Times.	
Keep "news" category in merged website		how people in india read newspapers
Add item in merged website		
• weather report on The Time	like weather report on the Time	
• footer at Sunday Times	like footer at Sunday times	
Categorise, compare, merge and change		
• keep review and		"People always want to

opinion as a whole in one category		hear advice about what to see in newspaper.”
• put money under Business	don’t think money is necessary in main category	“when people see a magazine, they see current rate or something”
• put culture, style and travel together under a category		”Culture and style are always together and when you travel, the culture and style will change”

Table 4: The data analysis of the observation from Participant 4 (female) from an embedded systems background.

Appendix d. Post-Interview analysis

Strategy	Context	Evidence
Participant strategy is to find the best feature of both websites and keep them.	previous experience of shopping	Mannequins wearing different kind of dresses, now I want to have like cap from one and scarf from other Mannequins, its like the customer perspective or users perspective

Table 5: The Narrative analysis of Participant 1(male) from a communication engineering background.

Strategy	Context	Evidence
At first glance she wants to see the Structure of the website	first glance	"I see the main page first, I see the layout, the position of the logo, menu bar" Complicating action: "At first glance, I want to see something structural and then I see information in detail"
Find the most important thing on the website	Culture	"because I think the most important thing in the side will be in the left in the centre, " "because in European Country we read from left to right"
Find the strength of both websites and merge them	Personal preference and culture	"I like the idea about the image in Sunday times" "In my country we have common phrase that 'An image worth 1000 words'"

Table 6: The narrative analysis of Participant 2 (female) from an usability and design background.

Strategy	Context	Evidence
He sees the underlying structure of the website.	culture	"I mean like the underlying structure of the website is quite important,"
find similarity and dissimilarity between different categories	patterns and meanings	"I try to figure out which one is same and have different name"
He tries to find the purpose or function in various element	meaning	"and so i try to see which kind of purpose or function certain element have"

He merge similar	meaning	“and try to combine that are similar,”

Table 7: The narrative analysis of Participant 3 (male) from an usability and design background.

Strategy	**Context**	**Evidence**
keep important thing	previous experience	“yes becoz review and opinion is really important” “Its like if you buy a book, you will first ask how is the book, have you read it.”
keep both “review” and “opinion”	meaning	“People opinion and review are a bit similar but not the same”
merge related subcategories into a single	meaning	“becoz your lifestyle depends upon your culture and your style defines your lifestyle and according to

Table 8: The narrative analysis of Participant 4 (female) from an embedded systems background.

Appendix e. Interview questions

1. How do you like the two websites? Why?

(Let participants talk about his understanding about the website, by asking this question we want to know if they have preference and whether the preference will affect the way they compare and merge)

2. How do you compare them? Why?

(By asking this question we want to know what affect them when they compare: style of the website, content of the website, prev knowledge, past experience)

3.What was your strategy of merging the two websites? Why?

(By asking this question we want to know what factors affect them when they merge)

Appendix f. Consent for participants

Experiment 1 (Includes Participant observation and interview):
by Arindra Kumar Das <akdas@kth.se>
Weiwei Zhang <weiweizh@kth.se>
for II2202 Research Methodology and Scientific Writing

Participant
Name: .....

Websites to compare and merge:
The Times (http://www.thetimes.co.uk/tto/news/)
and
Sunday Times (http://www.thesundaytimes.co.uk/sto/)

Introduction:
You will have two tasks and you have plenty of time to do it. You will be shown two websites, and your task is to compare them and merge them. Feel free to tell us anything that appear in your mind when you are doing the task. You are also free to use any knowledge and any help (google, call a friend, or ask us).
And you can use any tools here to illustrate your idea of how to merge them.(You can talk about it, you can draw it, or you can use software to make it.)
Inform us when you finish the task.

The process will be recorded by cameras, voice recorder and a screen record software. The data collected will be used only for this research study and will be kept anonymous.
We will provide some refreshments too.

By signing it you have read the above and allow us to use the data.

You are free to collect a PDF version of our report, in that case please leave your email here. .....
Sign with date and place.

.....