Interpreting Genre Evolution on the Web:
Preliminary Results

Marina Santini
University of Brighton
Lewes Rd, Moulsecoomb Campus,
Brighton, BN2 4GJ (UK)
Marina.Santini@itri.brighton.ac.uk

Abstract

The study presented in this paper explores the current state of genre evolution on the web through web users' perception. More precisely, it explores the perception of genres when users are faced not only with prototypical genre exemplars but also with hybrid or individualized web pages, and interpret the subjects' perception in term of genre evolution Although this exploration is partial (23 labels to be assigned to 25 web pages), it offers an interesting section of the genre repertoire on the web. This study can be also seen as a confirmatory study, because it confirms that a number of recent web genres, unprecedented in the paper world (such as home page, FAQs, and blog) can be recognized by the subjects; others have not fully emerged and many web users are not familiar with their new genre labels; finally some web pages show a high level of ambiguity and web users largely disagree on assigning labels to them.

1 Introduction

The study presented in this paper has been designed to explore the current state of genre evolution on the web through web users' perception. The web can be interpreted, among other things, as a genre repertoire in evolution because there are still many genre labels which have not been consolidated and many web pages that cannot be sorted into a recognized and acknowledged genre. The interpretation of the web as a genre repertoire in evolution has been developed within a research project on automatic identification of genre in web pages (Santini, 2006b). This interpretation is an attempt to explain the high level of hybridism and individualization of many web pages, which result in classification intractability. The study reported in this paper shows that humans have the same problems as classification algorithms when it comes to less standardized and conventionalized web pages.

As it has often been pointed out (for example, cf. Kwasnik and Crowston, 2005), it is hard to pin down the concept of genre from a single perspective or to find an agreed definition of what genre is. This lack is also experienced in the more restricted world of non-literary or non-fictional genres, such as professional or instrumental genres, where the variation due to personal style is less pronounced than in literary genres. In particular, scholars working with practical genres focus upon a specific environment. For instance Swales (1990) develops his notion of genre in academic and research settings, Bathia (1993) and Trosborg (2000) in professional settings, Yates and Orlikowsky (1992) within organizational communication. Despite the lack of an agreed theoretical notion, genre is a well-established term (cf. Karlgren, 2004), intuitively understood in its vagueness. Classifying documents by genre is a common operation that humans perform with more or less effort.

Genres can be seen as "artifacts", i.e. cultural objects created to meet and streamline communicative needs. These cultural objects represent the role that a certain type of documents plays in an environment. Each genre shows a set of standardized or conventional characteristics that makes it recognizable among

others, and this kind of identity raises specific expectations in the recipients, despite the fuzziness of genre labels (cf. Santini, 2005). Being cultural objects, showing common conventions and raising similar expectations are unifying traits. Together with these, there is a number of separating traits, such as hybridism, individualization, and evolution. In fact, genres are not mutually exclusive and different genres can be merged in a single document, generating hybrid forms. Genres are based on conventions, but allow a certain freedom of variation and consequently can be individualized.

Being *artifacts*, sharing *conventions* and *expectations*, showing *hybridism* and *individualization*, and undergoing *evolution* are important traits characterizing all sorts of genres. More precisely, genres can be defined as *cultural artifacts*, i.e. objects linked to a culture, a society or a community, bearing standardized features (conventions) but leaving space for creativity (individualization). On the one hand, standardized and recurrent features induce predictable expectations in the receivers. On the other hand, the freedom allowed by creativity allow genres to change, evolve, and be created to meet new needs (genre evolution), especially under the impulse of a new communication medium. While the change is still ongoing, i.e. before a modified genre is redefined, or a new genre is identified with a new name, documents show mixed forms and functions (genre hybridism).

This view of genre is flexible enough to encompass not only paper genres (both literary and practical genres), but also digital genres and, more specifically, web genres, such as the personal home page. The personal home page has no evident antecedent in the paper world (cf. Dillon and Gushrowski, 2000). It sprang up on the web as a new cultural object servicing the community of web users. When browsing a personal home page, web users expect a blend of standardized information (self-narration, personal interests, contact details, and often pictures related to one's life) and personal touch.

Another important thing to notice is that before genre conventions become fully standardized, genres do not have an official name. A genre name becomes acknowledged when the genre itself has a role and a communicative function in a community or society (Görlach, 2004: 9). Before this acknowledgement, a genre shows hybrid or individualized forms, and undefined functions.

For example, before 1998 web logs (or blogs) were already present on the Web, but they were not identified as a genre. They were just "web pages", with similar characteristics and functions. In 1999, suddenly a community sprang up using this new genre (Blood, 2000). Only at this point, the genre label "web log" or "blog" started spreading and being recognized.

Genre hybridism and individualization are evident on the web, and play an important role in the change and the creation of new genres. In fact, web pages – which can be considered as a new kind of document, much more unpredictable and customized than paper documents (Santini, 2006a) – are often hybrid because of intra-genre and inter-genre variations. They are also highly individualized because of the creative freedom provided by HTML or XML tags (the building blocks of web pages) or programming languages such as Javascript. On the web, new genres are constantly added (blogs, clogs, eshops, wikis, etc.) and traditional genres are adapted or updated in order to include more or different functionalities (online front pages, ezines, net ads, etc.). Genres such as emails, newsletters, search pages, eshops, etc. were a futuristic prophecy only 10 or 15 years ago, while today they belong to the normal life of a web user. Presumably, other genres will soon be added to meet new communicative needs brought about by new technologies.

As any other evolutions, also genre evolution proceeds along the axis of time. It is a diachronic process. There must be a 'before' and an 'after'. What often hallmarks a 'before' and an 'after' is the introduction of a new communication medium within a culture, a society, a community. The added value of studying genres on the web (a new medium) is represented by the possibility of following the development of genres and genres repertoires *live*, i.e. while it is taking place, and not *a posteriori*. That is, on the web we can capture synchronically a diachronic process. From a synchronic point of view, the genre repertoire is a continuum, where there are three forces interacting: what we bring from the past (reproduced genres), what is new or adapted to the new environment (novel genres and adapted genres), what is going to emerge and is not fully formed yet (emerging genres).

This view of genre evolution complements previous studies on the same subject (cf. Crowston and Williams, 1997; Shepherd and Watters, 1998; Kwasnik and Crowston, 2005). The main contribution of the synchronic

continuum is that an additional force has been acknowledged to take part in the evolution process, i.e. emerging genres. Emerging genres are those that are not fully standardized, that are still in formation and for which a genre label has not been created or have a label which is still opaque to the majority of users. Currently many web pages are in this phase of evolution, showing a high level of hybridism or individualization. We suggest that the subjects' perception of these web pages can be interpreted in term of genre evolution.

The study reported in this paper provides a snapshot of the current state of the genre repertoire of web pages seen through the perception of web users. Although this view is partial (23 labels to be assigned to 25 web pages), it offers an interesting section of the genre repertoire on the web. This study can be also seen as a confirmatory study, because it confirms that a number of recent web genres, unprecedented in the paper world (such as home page, FAQs, blog) can be recognized by the subjects; others have not fully emerged and many web users are not familiar with their new genre labels; finally some web pages show a high level of ambiguity and web users largerly disagree on assigning labels to them.

The article is organized as follows: Section 2 presents a short overview of previous work; Section 3 describes the web study and presents preliminary results; in Section 4 some conclusions are drawn.

2 Previous Work

No studies have been carried out so far on users' perception of a genre repertoire in transition.

Crowston and Williams (1997) were the first who reported on the genre repertoire on the web. They identified 48 reproduced and emergent genres in a sample of about 1,000 web pages.

A few user studies were carried out with the more pragmatic approach of exploring the usefulness of genre to improve web searches and defining a genre palette appropriate for this purpose. The most comprehensive study related to genre effectiveness for web searching is recent. Rosso (2005) carried out a series of four linked experiments, all based on human subjects. Quite surprisingly, the conclusion drawn by the author was that genre-annotated search results produced no significant improvement in participants' ability to make more consistent or faster assessment on the relevance of search results (Rosso, 2005: 133-179). In fact, only 17 of 32 participants reported noticing the genre label (Rosso, 2005: 176). Most probably, as pointed out by the author, this outcome was influenced by the difficulty and complexity of the task, together with the limitations of the setting (Rosso, 2005: 170-172).

Rosso's attempt to assess the relevance of search results including genre labels was almost unique. All other studies with web users, in contrast, did not provide any assessment of how well genres improved a web search. These studies are more like surveys on users' preferences in terms of useful non-topical categories that can help restrict web searches.

Along this line, Meyer zu Eissen and Stein (2004) built a genre palette for the web using two criteria: usability and feasibility. Their user study was based on a questionnaire where they asked about search engine use, usefulness of genre classification, and usefulness of genre classes. Interestingly, the authors note that one of the inherent problems of genre classification is that "even humans are not able to consistently specify the genre of a given page" because web pages have different functions, i.e. they might be hybrid forms, as in the case of product information sites that are combined with a shopping interface.

Roussinov et al. (2001) carried out a exploratory study of web users in order to identify what genres they most/least frequently come in contact with, and what genres most/least address their information needs. In their study, carried out in 2000, 116 different genres were identified, but not all web pages could be classified.

Karlgren (2000: 99 ff.), a pioneer in building a genre palette, tried to collect genres that were both consistent with what users expect as well as conveniently computable. He sent around a questionnaire where the core question was: "What genres do you feel you find on the WWW?". He ended up with a palette of 11 genres. One frequent comment by the respondents was that the genres in the palette were not mutually exclusive, in other words they showed some level of hybridism.

Very informative in many respects, these studies have in common the practical aim of improving web searches. This might explain why they overlook difficult issues such as the hybridism or the individualization of many web pages, which are nonetheless perceived by the subjects. The authors must necessarily focus on

unambiguous exemplars, showing clear-cut conventions and expectations.

The present study, on the other hand, explores the perception of genres when users are faced not only with prototypical genre exemplars but also with hybrid or individualized web pages, and interpret the subjects' perception in term of genre evolution.

3   Web Study

The study described in this section was web-based. It was uploaded on to one of the servers at University of Brighton at the end of February 2005, and kept online for one month.

The study is based on participants who volunteered within the University of Brighton (UK), University of Sussex (UK), Dalhousie University (Canada), Syracuse University (USA), plus other academics (interested in genre-related issues) in other universities and research institutes in Europe. Potential participants were sent an email containing the URL of the study on the web.

3.1   Population and Sample: Academic Environment

Genre recognition and acknowledgement is based on elements like education, culture, community, and society. The academic population on which the study is built upon has three elements in common:

- it is a medium-high educated population (from administrative people to students and professors);

- it is very used to computer-meditated communication;

- it is familiar with the Web.

3.2   Web Pages and Web Genres

Web pages were chosen by the author of this paper from the *live* Web and from the SPIRIT collection of web pages (Joho and Sanderson, 2004). Three typologies of web genres and web pages were hypothesized for the selection and for the study (the web pages included in the study are available, together with their URLs, at http://www.nltg.brighton.ac.uk/home/Marina.Santini/:

1) Easy web genres:
```
1.    eshop (web_page_01)
2.    personal home page (web_page_02)
3.    front page (web_page_04)
4.    search page (web_page_05)
5.    corporate home page (web_page_11)
```

```
6.    FAQs (web_page_12, the word "FAQs"
      was deleted from the heading)
7.    splash screen (web_page_24)
8.    net ad (web_page_2)
```

2) Ambiguous web genres:
```
9.    email (web_page_03, because of the
      format and the granularity: email vs.
      mailing list)
10.   sitemap (web_page_06, the words
      "sitemap" and "hotlist" were deleted
      from the heading)
11.   hotlist (web_page_15,the word
      "hotlist" was deleted from the
      heading),
12.   academic personal home page
      (web_page_08)
13.   about page (web_page_10)
14.   organizational home page
      (web_page_14)
15.   blog (web_page_07)
16.   clog (web_page_16,blog and clog could
      be swapped in their interpretation)
17.   search by multiple fields
      (web_page_17)
18.   online form (web_page_10, online
      forms and search by multiple field
      can appear very similar)
19.   newsletter (web_page_19, which was
      presented truncated),
20.   howto page (web_page_20)
21.   online tutorial (web_page_22, online
      tutorial is a super-genre of howto
      pages)
```

3) Difficult web pages:
```
22.   ezine cover (web_page_13)
23.   "Adirondack Orienteering Klub"
      (web_page_18, the author could not
      find a genre for it)
24.   CitiDex (web_page_21, the author
      could not find a genre for it)
25.   Collimating Lens Holder (web_page_23,
      the author could not find a genre for
      it)
```

The expectation was that easy web genres would collect the highest rate of agreement, ambiguous web genres would receive a lower agreement rate, while difficult web pages were expected to be the most controversial in users' perception.

The term "genre" was never mentioned in the whole study in order not to influence or confuse the participants. The goal of the study was not declared either because the idea was to ask for a genre classification of web pages implicitly and study the reactions. Participants were simply told to assign "labels" to web page "types".

3.3   Participants' Task and Sample Size

The task of participants was straightforward. They had to go through 25 screenshots of web pages and assign one of the 23 labels to each of them.

The total number of users who started the experiment was 198. 135 participants went

through the whole study and provided valid responses for the experiment.

3.4 Results

Currently, there is no standard test largely agreed upon that can be used for experiments where subjects can make choices from a large number of categories (23 labels) for a large number of objects (25 web pages). In the following paragraphs some views and interpretations of the data are presented, namely raw counts and percentages, Fisher's exact test, and adjusted residuals.

Raw Counts and Percentages. A view on the data is offered in Table 2, which shows the number of subjects assigning a particular label to a particular web page and the percentage of the most voted label. For example, the label *eshop* (8th row) was assigned to WP1[1] (first column) by 119 subjects (highlighted cell), which corresponds to 88.15% (bottom row). Four subjects thought that WP1 was a *corporate home pages* (around 2.9%), seven selected *net ad* (around 5%), one subject chose *front page* (around 0.7%), one *hotlist*, one *did not know*, two added a new label for it (around 1.4%).

Three ranges of agreement can be identified out of this table. The top range includes web pages with a percentage of agreement above 80%; the middle range groups web pages with an agreement between 79% and 50%; finally the bottom range contains web pages with an agreement between 49 % and 20%. Table 1 lists the web pages by percentage of agreement.

From these ranges a first conclusion can be drawn. According to the ranges shown in Table 1, participants show the highest agreement on what we selected as "easy web genres", except in three cases: front pages, net ad and splash screen, which seem among the least agreed upon (see bottom range). The middle range includes most of the ambiguous web genres together with ezine, which was deemed to be difficult by the author. The bottom range includes the rest of the ambiguous genres, together with other difficult web pages and three web pages from the top range, webpage_type_04 (front page), webpage_type_24 (splash screen) and webpage_type_25 (net ad).

We have now a first picture of users' perception of some web pages in relation to some web genre labels. The hypothesized genre recognition pattern was mostly confirmed in the top range, but slightly reshuffled in the middle and bottom ranges. Figure 1 shows the charted percentages.

Fisher's Exact Test. The percentages at the bottom row in Table 2 can be interpreted in terms of conditional distribution on the most voted label (response variable) per web page type (explanatory variable). In other words, they refer to the sample distribution of most voted labels, *conditional* to the web page type. In terms of association, this means that the distribution of the response variable (the label) changes with the value of the explanatory variable (the web page type) if the two variables are related. Table 2 suggests the existence of an association or correlation between the label and the web page to which this label was assigned. But as Table 2 refers to the sample rather than the population, it provides evidence but not the final answer to whether labels and web page types are associated in the way suggested by the percentages. In order to see if it is plausible that labels and web page types are associated in the population, Fisher's exact test can be calculated. The value returned for this test by SPSS is 9292.275, which is large enough to reject the hypothesis that labels and web page are independent[2]. This statistically significant association shows that the web pages chosen by the author to represent some web genres mostly map the subjects' perception of these web pages. It also shows that many genre labels are acknowledged by the users and are consistently associated to web pages.

Adjusted Residuals: A test statistic, such as Fisher's exact test, and statistical significance summarize the strength of evidence against the null hypothesis of independence, but does not indicate how many and which cells deviate greatly from this hypothesis. Residuals, i.e. the differences between expected and observed cell frequencies can help in this task. In particular, adjusted residuals can indicate if the cell counts are significantly different from what independence predicts. A large adjusted residual provides evidence against independence of a cell. As Table 3 mostly maps Table 2, a significant association between genre labels and web page types on the cells containing the most voted labels is then confirmed.

---

[1] WP1, WP2, WP3, etc. are short form of webpage_01, webpage_02, webpage_03, etc.

[2] The larger the value, the greater the evidence against the null hypothesis of independence.

3.5 Discussion

The original impression that there were different degree of perception of genres of web pages was confirmed by these preliminary results. Also the rough distinction into three levels of genre awareness (easy, ambiguous and difficult) was confirmed. Three ranges of perception came out clearly from percentages, but the distribution of the web pages into these three ranges is slightly different from what was expected.

The general view of the results (Fisher's test) reveals that there is a significant association between the 25 web pages and the 23 labels. The analysis of adjusted residuals support this interpretation.

The agreement among subjects on the label to assign to a particular web pages can be divided into three levels.

At the first level, which can be interpreted as the highest perception of web genres, there are web pages labelled as personal home page (webpage_type_02), eshop (webpage_type_01), corporate home page (webpage_type_11), FAQs (webpage_type_12), and search pages (webpage_type_05). We can define these labels as stable web genres.

At a middle level of perception, there are web genres still emerging. Most of the labels are fairly novel (ezine, clog, blog, about, how to), sometimes not entirely transparent, and some of them are specialized (academic home page, organizational home page, online tutorial). Probably the textual conventions of these genres are not entirely standardized yet and can cause oscillation in users' perception. This level offers the most interesting view on a genre repertoire which is moving and evolving and it is not consolidated yet.

The bottom range shows a blurred level of perception for different reasons. For some genres such as email and newsletter, the presentation in form of screenshots was not ideal. Subjects could not navigate through the web page and they could not resolve the level of granularity. For instance, for webpage_type_03 (the web page selected by the author to represent an email), 66 subjects chose email, but 34 subjects preferred to add a new label for it and 20 thought it was an about page. Surprisingly, labels such as splash screen and front page for webpage_type_04 and webpage_type_24 were not favoured by the respondents who preferred to add their own labels in many cases. For webpage_type_06, subjects preferred the label search page instead

of sitemap. Another interesting case is net ad (webpage_type_25), which was often assessed as eshop, probably because the concept of advertising and selling are closely related. The most opaque label seems to be hotlist (webpage_type_15) because most subject preferred to add their own label. Three of the four web pages that were classified by the author as "I don't know" belong to this level of perception. While webpage_type_21 fell into the middle range because most of the subjects perceive it as a search page, the genre perception or interpretation of webpage_type_17, webpage_type_18, and webpage_type_23 is not so straightforward. For instance, webpage_type_17 was assessed as online form (57 subjects), search page (26 subjects), an eshop (26 subjects) and probably it is has all these functions at the same.

4 Conclusions and Future Work

The study shows a composite picture of the perception of the genre repertoire on the Web. This picture focuses on recent genres only, overlooking those more based on paper genres because, in our opinion, this hot area can reveal more about the dynamics behind genre evolution.

Preliminary findings coming out from this study confirm the initial hypothesis and show that users' perception can be divided into three ranges. These three ranges can be interpreted in terms of genre evolution: high perception for the most stable and acknowledge genres; medium perception for emerging genres, not fully acknowledged by the majority or still unstable, and finally low perception for the highly ambiguous genres (for different reasons). Some of the new web genres can be unambiguously perceived (for example, personal home page, eshop, corporate home page, FAQs and search page).

Web users can also handle a certain degree of granularity, for example by distinguishing a personal home page from a corporate home page, but the boundary between academic home pages and organizational home pages is still too fuzzy for them.

The approach to the web as a genre repertoire in evolution and these preliminary findings can turn out to be useful when building web genre palettes or when designing new genre identification experiments.

Future work includes the computation of agreement coefficients. K statistic is largely used

but still controversial and mostly used for measuring the agreement of two or three raters. Two new interesting measures to assess users' recognition of web page genres were used by Rosso (2005: 109 ff.), but their full interpretation is still under study. The challenging follow up of these preliminary results is to find an objective coefficient of agreement applicable for 135 raters that can choose among 23 categories to classify 25 objects.

References

Bathia Vijay (1993), *Analysing Genre. Language Use in Professional Settings*, Longman, London-NY.

Blood R. (2000),Weblogs: A History and Perspective, http://www.rebeccablood.net/essays/weblog_history.htm

Crowston K. and Williams M. (1997), Reproduced and Emergent Genres of Communication on the World-Wide Web, *Proc. 30th Hawaii International Conference on System Sciences*.

Dillon A. and Gushrowski B. (2000), Genres and the Web: is the personal home page the first uniquely digital genre?, *Journal of the American Society for Information Science*, 51, 2.

Görlach M. (2004), *Text Types and the History of English*, Mouton de Gruyter, Berlin-NY.

Joho H. and Sanderson M. (2004), The SPIRIT collection: an overview of a large web collection, *SIGIR Forum*, December 2004 (Vol. 38, N. 2)

Karlgren J. (2000), *Stylistic Experiments for Information Retrieval*, Thesis submitted for the degree of PhD, Stockholm University, Sweden.

Karlgren J. (2004), The Wheres and Whyfores for Studying Textual Genre Computationally, *Papers from the AAAI Fall Symposium*.

Kwasnik, B. and Crowston, K. (2005), Genres of digital documents: Introduction to the special issue, *Information, Technology & People*, 18(2), 76–88.

Meyer zu Eissen S. and Stein B. (2004), Genre Classification of Web Pages: User Study and Feasibility Analysis, in Biundo S., Fruhwirth T. and Palm G. (eds.), *KI 2004: Advances in Artificial Intelligence*, Springer, Berlin-Hedelberg-NY, p. 256-269.

Rosso M. (2005), *Using Genre to Improve Web Search*, Thesis submitted for the degree of PhD, University of North Carolina at Chapel Hill, USA.

Roussinov D., Crowston K., Nilan M., Kwasnik B., Cai J., Liu X. (2001), Genre Based Navigation on the Web, *Proc. 34th Hawaii International Conference on System Sciences*.

Santini M. (2005), Genres In Formation? An Exploratory Study of Web Pages using Cluster Analysis, *Proc. CLUK 05*.

Santini M. (2006a), Web pages, text types, and linguistic features: Some issues, *ICAME Journal*, Vol. 30.

Santini M. (2006b), Automatic Identification of Genres in Web Pages, *forthcoming*.

Shepherd M. and Watters C. (1998), The Evolution of Cybergenre, *Proc. 31st Hawaii International Conference on System Sciences*.

Swales J. (1990), *Genre Analysis. English in academic and research settings*, Cambridge University Press, Cambridge.

Trosborg Ann (ed.) (2000), *Analysing Professional Genres*, John Benjamins Publishing Company, Amsterdam-Philadelphia.

Yates J., and Orlikowski W. (1992), Genres of organizational communication: A structural approach to studying communications and media, *Academy of Management Review*, Vol. 17, 2, 229-326.

Appendix

Ranges		Web page type and related web genre suggested by the author
Top: Above 80%	5	webpage_type_01 (eshop), webpage_type_02 (personal_hp), webpage_type_05 (search_page), webpage_type_11 (corporate_hp), webpage_type_12 (faqs)
Middle: From 79% to 50%	10	webpage_type_07 (blog), webpage_type_08 (academic_hp), webpage_type_09 (online_form) webpage_type_10 (about_page) webpage_type_13 (ezine) webpage_type_14 (organiz_hp) webpage_type_16 (clog) webpage_type_20 (howto) webpage_type_21 (dontknow) webpage_type_22 (tutorial)
Bottom: From 49% to 20%	10	webpage_type_03 (email) webpage_type_04 (frontpage) webpage_type_06 (sitemap) webpage_type_15 (hotlist) webpage_type_17 (dontknow) webpage_type_18 (dontknow) webpage_type_19 (newsletter) webpage_type_23 (dontknow) webpage_type_24 (splashscreen) webpage_type_25 (netad)

Table 1. Ranges of percentages

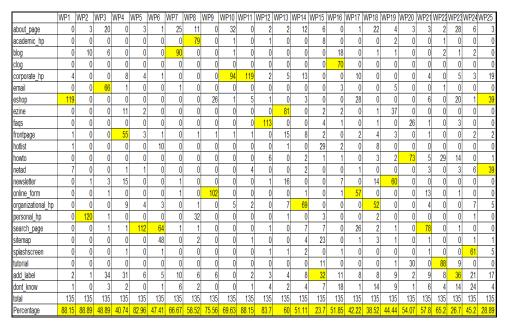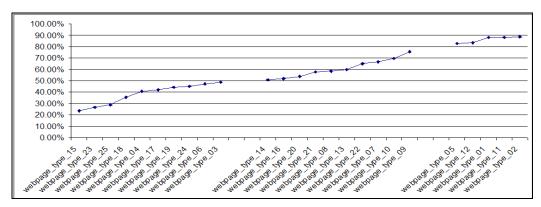|	WP1	WP2	WP3	WP4	WP5	WP6	WP7	WP8	WP9	WP10	WP11	WP12	WP13	WP14	WP15	WP16	WP17	WP18	WP19	WP20	WP21	WP22	WP23	WP24	WP25
about_page	0	3	20	0	3	1	25	11	0	32	0	2	2	12	6	0	1	22	4	3	3	2	28	6	3
academic_hp	0	0	0	0	0	0	0	79	0	1	0	1	0	0	8	0	0	2	0	0	1	0	0	0	0
blog	0	10	6	0	0	0	90	0	1	0	1	0	0	0	0	0	18	0	1	1	0	0	2	1	2
clog	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	70	0	0	0	0	0	0	0	0	0
corporate_hp	4	0	0	8	4	1	0	0	0	94	119	2	5	13	0	0	10	0	0	0	4	0	5	3	19
email	0	0	66	1	0	0	1	0	0	0	0	0	0	0	0	3	0	0	5	0	0	1	0	0	0
eshop	119	0	0	0	0	0	0	0	0	26	1	5	1	0	3	0	0	28	0	0	0	6	0	20	39
ezine	0	0	0	11	2	0	0	0	0	0	0	0	81	0	2	2	0	1	37	0	0	0	0	0	0
faqs	0	0	0	0	0	0	0	0	0	0	0	0	113	0	4	1	0	1	0	26	1	0	3	0	0
frontpage	1	0	0	55	3	1	0	1	1	1	1	0	15	8	2	0	2	4	3	0	1	0	0	2	2
hotlist	1	0	0	0	0	10	0	0	0	0	0	0	1	0	29	2	0	8	0	0	0	0	0	0	0
howto	0	0	0	0	0	0	0	0	0	0	0	6	0	2	1	1	0	3	2	73	5	29	14	0	1
netad	7	0	0	1	1	0	0	0	0	0	4	0	0	2	0	0	1	0	0	0	3	0	3	6	39
newsletter	0	1	3	15	0	0	1	0	0	0	0	1	16	0	0	7	0	14	60	0	0	0	0	0	0
online_form	0	0	1	0	0	0	1	0	102	0	0	0	0	1	0	1	57	0	0	0	0	13	0	1	0
organizational_hp	0	0	1	9	4	3	0	1	0	5	2	0	7	69	0	0	0	52	0	0	4	0	0	7	5
personal_hp	0	120	1	0	0	0	0	32	0	0	0	0	1	0	3	0	0	2	0	0	0	0	0	1	0
search_page	0	0	1	1	112	64	1	1	0	0	0	1	0	7	7	0	26	2	1	0	78	0	1	0	0
sitemap	0	0	0	0	0	48	0	2	0	0	1	0	0	4	23	0	1	3	1	0	1	0	0	1	1
splashscreen	0	0	0	1	0	1	0	0	0	0	0	1	1	2	0	1	0	0	0	0	1	0	0	61	5
tutorial	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	1	30	0	88	9	0	0
add_label	2	1	34	31	6	5	10	6	6	0	2	3	4	8	32	11	8	8	9	2	9	8	36	21	17
dont_know	1	0	3	2	0	1	6	2	0	0	1	4	2	4	7	18	1	14	9	1	6	4	14	24	4
total	135	135	135	135	135	135	135	135	135	135	135	135	135	135	135	135	135	135	135	135	135	135	135	135	135
Percentage	88.15	88.89	48.89	40.74	82.96	47.41	66.67	58.52	75.56	69.63	88.15	83.7	60	51.11	23.7	51.85	42.22	38.52	44.44	54.07	57.8	65.2	26.7	45.2	28.89

Table 2. Users' assignment and percentages

Figure 1. Charted percentages

	WP01	WP02	WP03	WP04	WP05	WP06	WP07	WP08	WP09	WP10	WP11	WP12	WP13	WP14	WP15	WP16	WP19	WP20	WP22	WP24	WP25
about_page	-2.7	-1.4	5.6	-2.7	-1.4	-2.2	7.7	1.9	-2.7	10.6	-2.7	-1.8	-1.8	2.3	-0.2	-2.7	-1.0	-1.4	-1.8	-0.2	-1.4
academic_hp	-2.2	-2.2	-2.2	-2.2	-2.2	-2.2	-2.2	37.1	-2.2	-1.7	-2.2	-1.7	-2.2	-2.2	-2.2	-2.2	-2.2	-1.2	-2.2	-1.7	-2.2
add_label	-2.8	-3.1	7.8	6.8	-1.5	-1.8	-0.1	-1.5	-1.5	-3.4	-2.8	-2.4	-2.1	-0.8	7.2	0.2	-0.5	-2.8	-0.8	3.5	2.2
blog	-2.6	1.6	-0.1	-2.6	-2.6	-2.6	35.3	-2.6	-2.6	-2.2	-2.6	-2.6	-2.6	-2.6	-2.6	5.0	-2.2	-2.6	-1.8	-1.8	-2.6
clog	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	-1.9	37.9	-1.9	-1.9	-1.9	-1.9	-1.9
corporate_hp	-2.7	-3.9	-3.9	-1.5	-2.7	-3.6	-3.9	-3.9	-3.9	24.3	31.8	-3.3	-2.4	0.0	-3.9	-3.9	-3.9	-3.9	-3.9	-3.0	1.8
dontknow	-1.7	-2.2	-0.7	-1.2	-2.2	-1.7	0.8	-1.2	-2.2	-2.2	-1.7	-0.2	-1.2	-0.2	1.3	6.7	2.3	-1.7	-0.2	9.7	-0.2
email	-2.0	-2.0	33.8	-1.4	-2.0	-2.0	-1.4	-2.0	-2.0	-2.0	-2.0	-2.0	-2.0	-2.0	-2.0	-0.4	0.7	-2.0	-1.4	-2.0	-2.0
eshop	38.2	-3.2	-3.2	-3.2	-3.2	-3.2	-3.2	-3.2	5.8	-2.9	-1.5	-2.9	-3.2	-2.2	-3.2	-3.2	-3.2	-3.2	-3.2	-2.9	10.4
ezine	-2.7	-2.7	-2.7	1.9	-1.8	-2.7	-2.7	-2.7	-2.7	-2.7	-2.7	-2.7	30.9	-2.7	-1.8	-1.8	12.7	-2.7	-2.7	-2.7	-2.7
faqs	-2.8	-2.8	-2.8	-2.8	-2.8	-2.8	-2.8	-2.8	-2.8	-2.8	-2.8	42.6	-2.8	-2.8	-1.1	-2.4	-2.8	7.7	-2.8	-2.8	-2.8
frontpage	-1.7	-2.2	-2.2	24.6	-1.7	-1.7	-2.2	-1.7	-1.7	-1.7	-1.7	-2.2	-1.7	1.7	-1.3	-2.2	-0.8	-2.2	-2.2	-1.3	-1.3
hotlist	-0.8	-1.5	-1.5	-1.5	-1.5	5.7	-1.5	-1.5	-1.5	-1.5	-1.5	-1.5	-0.8	-1.5	19.4	0.0	-1.5	-1.5	-1.5	-1.5	-1.5
howto	-2.4	-2.4	-2.4	-2.4	-2.4	-2.4	-2.4	-2.4	-2.4	-2.4	-2.4	0.2	-2.4	-1.6	-2.0	-2.0	-1.6	30.2	10.5	-2.4	-2.0
netad	2.5	-1.8	-1.8	-1.1	-1.8	-1.8	-1.8	-1.8	-1.8	-1.8	0.7	-1.8	-1.8	-0.5	-1.8	-1.8	-1.8	-1.8	-1.8	1.9	22.1
newsletter	-2.3	-1.9	-0.9	4.7	-2.3	-2.3	-1.9	-2.3	-2.3	-2.3	-2.3	-1.9	5.2	-2.3	-2.3	1.0	25.8	-2.3	-2.3	-2.3	-2.3
online_form	-2.3	-2.3	-1.9	-2.3	-2.3	-2.3	-1.9	-2.3	45.1	-2.3	-2.3	-2.3	-2.3	-1.9	-2.3	-1.9	-2.3	-2.3	-2.3	-2.3	-2.3
organizationa	-2.4	-2.4	-2.4	1.7	-0.6	-1.1	-2.4	-2.0	-2.4	-2.0	-0.2	-1.5	-2.4	28.8	-2.4	-2.4	-2.4	-2.4	-2.4	0.8	-0.2
personal_hp	-2.9	43.2	-2.5	-2.9	-2.9	-2.9	-2.9	9.4	-2.9	-2.9	-2.9	-2.9	-2.5	-2.9	-1.7	-2.9	-2.9	-2.9	-2.9	-2.9	-2.9
search_page	-3.2	-3.2	-2.9	-2.9	35.7	19.0	-2.9	-2.9	-3.2	-3.2	-3.2	-2.9	-3.2	-0.8	-0.8	-3.2	-2.9	-3.2	-3.2	-3.2	-3.2
sitemap	-2.0	-2.0	-2.0	-2.0	-2.0	23.4	-2.0	-1.0	-2.0	-2.0	-1.5	-2.0	-2.0	0.1	10.1	-2.0	-1.5	-2.0	-2.0	-1.5	-1.5
splashscreen	-1.9	-1.9	-1.9	-1.4	-1.9	-1.4	-1.9	-1.9	-1.9	-1.9	-1.9	-1.4	-1.4	-0.8	-1.9	-1.4	-1.9	-1.9	-1.9	32.0	0.8
tutorial	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	-2.6	2.0	-2.6	-2.2	10.0	34.5	-2.6	-2.6	-2.6

Table 3. Adjusted residuals