Elastic lists for facet browsers

Moritz Stefaner, Boris Müller
Interaction Design Lab
University of Applied Sciences Potsdam
{moritz.stefaner, boris.mueller}@fh-potsdam.de

Abstract

In the context of current web and personal information management developments, we argue that facet browsing is an increasingly important interface paradigm. However, current implementations neglect two important aspects of metadata distributions: the relative proportions of metadata occurrences and the unusualness of this proportion compared to a global profile. Based on focus & context visualization techniques, we enhance facet browser user interfaces with "elastic lists" to make the resulting weighted metadata profiles visually accessible and navigable. The principle is currently developed and tested in several domains.

1 Trends in information management

1.1 Microcontent

Information presentation, storage and communication has been changed considerably by digital technologies. One recent trend is especially remarkable: Information items tend to get much shorter. This is not only an effect of the technologies used to publish and communicate information (such as blogging software, cell phones, email clients) but also the consumption behavior of the users and the according social practices. [3]

Usually, these minimal information items are referred to as microcontent. While Jacob Nielsen's original definition [10] focussed the production of easily skimmable text items based on the "inverted pyramid style", recent web developments lead to a wider understanding of the term. Dr. Arnaud Leene postulates several properties qualifying a digital information item as microcontent: Focussed, self–contained, indivisible, structured, adressable[9]. This qualifies also business cards, video clips or cooking recipes as microcontent and matches current web publishing and usage practices better than the original definition.

If represented in XHTML, microcontent can be structured by using microformats[1]. They allow an easy, machine–readable mark–up of microntent items such as business cards, event announcements or reviews in a standardized way. The value of a metadata field or a microformat component, however, is a primitive type such as string or number and hence establish only flat additional metadata information.

1.2 Web feeds

Additionally, web feeds introduced a new information delivery paradigm to the Web: Instead of actively accessing web pages of interest, web feeds allow users to subscribe to frequently updated contents. To consume web feeds, usually, a dedicated feed reader application is needed, but recent browser versions also support direct display and subscription of feeds. Originally used for news teasers pointing to the original stories, web feeds are increasingly used to

- deliver structured microcontent; e.g. weather information, blog posts, or media files (so called podcasts for audio files or vodcasts for video files).

- embed information from external sources into web pages or applications

- subscribe to queries on web applications (such as a subscription to a specific user's public bookmarks or photos taken at a specific place)

- transfer information between different devices, applications or web pages

From a metadata perspective, although web feeds represent a well-defined structured data format, the metadata contained in web feeds has simple nominal values or a standard date format. There is no agreed-upon mechanism to identify e.g. item authors across web feeds or refer to items in more complex information architectures such as domain ontologies.

1.3 Tagging

Tagging is the process of assigning freely chosen text labels ("tags") to objects, typically digital resources, for future navigation, filtering or search. Besides the semantics contained in the chosen tags, the act of tagging *per se* can already be used as a bookmarking or flagging gesture to contrast tagged from untagged content. Tagging is not exclusively descriptive: frequently also marker tags such as "toRead" or subjective judgements such as "cool" are used. Often, the time of the tagging activity is stored as additional metadata.

Collaborative tagging systems provide a framework for a user community to tag publicly available resources in a "socially translucent"[5] manner. These provide each user an awareness of both their individual tags as well as the tags and content that others contribute to the community. Only by providing "immediate self and social feedback" [15], stable, community–wide patterns in tag usage emerge over time[6]. The resulting multi–faceted, bottom–up organization is often referred is as folksonomy — a neologism based on the words "folk" and "taxonomy" [11]

Empirical analysis of emerging structures in tagging systems shows that although tagging is a rather unconstrained and possibly highly subjective activity, two stable patterns can be observed across data sets[4]:

- Re-use of tags seems to have a non-uniform distribution over time. In fact, tag production data can be modelled using a simple Yule-Simon process enhanced with a fat-tailed memory kernel, which makes recently used tags more probable to re-appear while leaving low access probability to tags used longer ago. This makes the temporal dynamics of tagging an important factor, which should strongly be taken into account when providing user interfaces in this domain.

- There is an interesting relation between high-rank tags and the high number of low-rank tags (the "long tail"): Also for very rarely used tags, usually one or more strongly correlated "parents" can be identified among the top-rank tags. On the other hand, top-rank tags seem to categorize the tag space quite well, although in a non-exclusive and non-trivial manner.

2 Consequences for the interface design of facet browsers

The discussed changes in information publishing and storage behavior and formats have some profound implications for designing interfaces in this domain:

- The resulting information architecture is typically flat, non-hierarchical and non-exclusive.
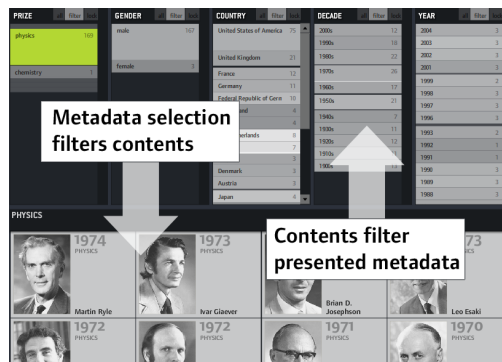

Figure 1. Facet browsing principle

- In contrast to catalogues or classification schemes, metadata values tend to follow a steep long tail distribution, where few values occur often, but a large number of entries is used rarely.

- Both by explicit rules as well as usage patterns, meaningful correlations between metadata fields emerge.

- Temporal dynamics and life-cycles of contents and metadata gain importance.

The exploration of dynamic taxonomies [13] with so–called facet browsers is often seen as a most promising candidates for "rich exploration of a domain across a variety of sources from a user-determined perspective"[8]. These make different aspects of the underlying data accessible in parallel. Selecting one of the values, and thus filtering the result set, restricts the available metadata values only to those ocurring in the results. Consequently, the user is visually guided through an iterative refinement process, never encountering situations with zero results. See Figure 1 for a screenshot with an illustration of the interaction principle.

However, current implementations often rely on a stable, hierarchically structured and well-designed metadata structure, also reflected in the common drill-down model of facet navigation. In the context of the described developments, the presentation layer of facet browsers has to be able to carve out the essentials of flat, yet correlated structures across multiple facets in order to enable navigation via metadata attributes. We argue that in order to make the long tail of metadata accessible, two important factors need to be used to visualize and pre-select shown values: relative proportions of metadata occurrences as well as the unusualness of these proportions compared to a wider context.
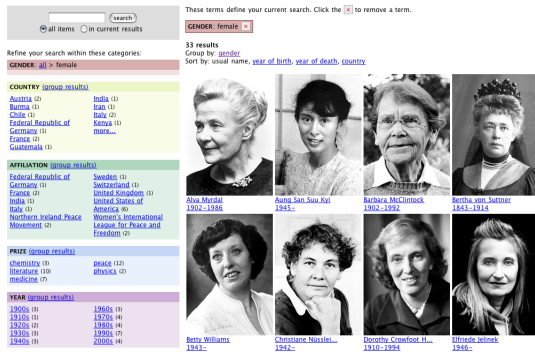
Figure 2. The Flamenco facet browser

Figure 3. Different states of an elastic list

3 Visualizing metadata profiles in elastic lists

3.1 Browsing metadata profiles

Central to our approach is the notion of *metadata profile*. If we define a context as a set of contents and their metadata values, a metadata profile expresses the characteristics of a given context in terms of its metadata distribution. In its simplest version, a metadata profile is represented as the set of occurring metadata values weighted by the number of occurrences.

The *global metadata profile* is the metadata profile for all available contents and hence represents the a priori distribution of metadata. A *local metadata profile* characterizes a subset of contents, such as a search result, the result of a filtering operation or a single content.

Regarding the classical Flamenco system[18] or other facet browser implementations [7],[2], it becomes evident that a certain standard for the presentation of metadata profiles has been established: Take, for instance, the Flamenco browser's "Nobel Prize winners" demonstration (see Figure 2): in initial navigation, the value "female" was selected from the attribute "gender". This filters the display of contents to ones matching this attribute value; in turn, all metadata attribute fields are restricted only to values occurring together with the selected attribute. On subsequent filtering steps, this makes it impossible to construct queries with an empty result set, which is commonly regarded as one of the biggest benefits of facet browsers. To generalize in the terminology introduced above, the facet navigation displays the local metadata profile for the selected context by employing a simple visual mapping: Only values with a weight greater than zero are presented, usually in a list and in visually uniform manner; often, the weight is presented as a number in parentheses.
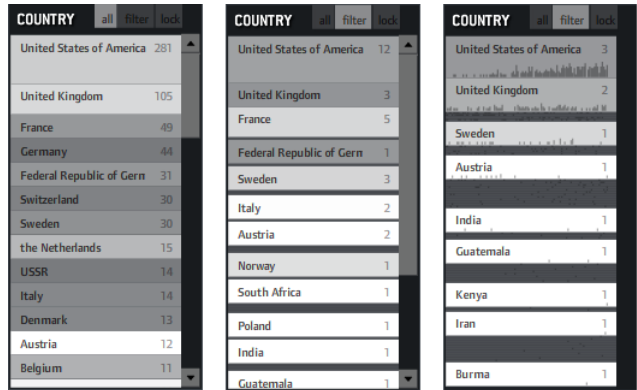
3.2 The elastic lists principle

We build upon the navigation principle of facetted browsing, but enhance the information presentation in so-called *elastic lists* with respect to the following features:

- **Visualize the weight proportions of attributes** In many situations, it is informative to immediately see which are the predominant values and which cover only a minor part of the data set.

- **Emphasize the characteristic values of a local profile** In order to understand what makes a data set special compared to the whole collection, it is helpful to indicate how the displayed proportions differ from the global distribution. In the Nobel Prize winners example, e.g. it would be informative to see that 35% percent of all female prize winners received a peace Nobel prize, while the global ratio is only 14,4%. This makes "peace prize" a characteristic attribute for the selected subset, which is not evident from a plain list presentation.

- **Animated filtering** For users of facet browsers, the sudden disappearance of list items after click is a common source for misconceptions and confusion. In our elastic list representation, transitions are animated smoothly and even filtered–out attribute values are still visible as flat lines. This makes the filtering process transparent to the user and allow easy localisation of the local metadata profile compared to the global profile.

Elastic lists follow the following principles: Items are presented in form of an ordered list. The size of an item indicates the proportion of items associated with the respective metadata value. The brightness of a list item indicates

the "unusualness" of an item weight in the given context. Two modes can be distinguished:

- In its initial state, an elastic lists display the global metadata profile. All items are visible. (see Figure 3a) The measure of unusualness is defined in terms of a trend measure — metadata values with recently rising activity are visually emphasized by a brighter background color. For ordinal data, such as time points, items are ordered descendingly; for nominal data either the trend measure or the weight in the global profile can serve as ordering principle.

- In their filtered state, elastic lists maintain the same order of items, but metadata attributes with a weight of zero (i.e. not occurring in the current context) are collapsed to a minimal visible height. All other metadata items are scaled according to their proportional weight[1]. A brighter color indicates that the proportional weight is significantly higher than compared to the global profile. (see Figure 3b)

Transitions between states are animated in order to facilitate an understand of the filtering process. Switching between "global" and "filtered" mode is possible at any time by using dedicated buttons. Any state of the elastic list can be frozen via the "lock" button to allow sequential exploration of the presented values without continuous transformation of the list.

Additionally, small bar charts (so–called "sparklines"[17]) indicating the temporal dynamics of the metadata value can be displayed (see Figure 3c). These represent a histogram of the ocurrence of the respective metadata value, with time points–in this case years–running from left to right.

3.3 Example: Nobel prize winners dataset

In order to make our approach directly comparable to other approaches, we implemented a demonstration based on the Nobel prize winners dataset used in Flamenco.[2] It should be noted, that the metadata structure used in the data set does not represent the previously described flat, yet interrelated structures induced by free–form tagging. Nevertheless, our visualization approach leads to interesting insights on the data set: When e.g. selecting the value

---

[1]Theoretically, the size of the list entries should correspond directly to their proportional weight. However, for usability reasons, each entry with a non-zero weight has been assigned a minimum height in order to make all entries of interest readable. Additionally, due to the often skewed distribution of values, a logarithmic transform on the weight is applied to dampen the influence of high weights.

[2]An interactive version is available at
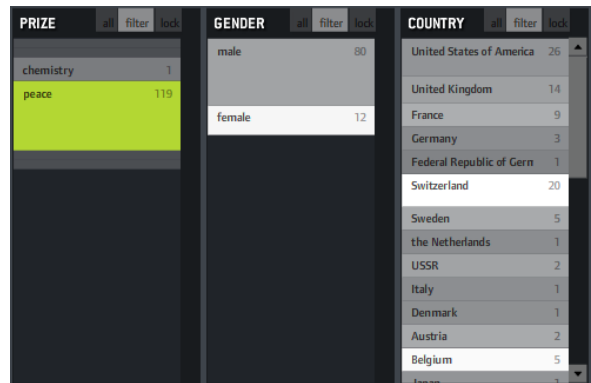http://well-formed-data.net/experiments/elastic_lists/.

Figure 4. For "peace" noble prizes, the metadata values "female", "Switzerland" and "Belgium" have an unusually high weight.

"peace" from the "prize" category, we can observe that although more men than women have achieved a peace nobel prize overall, the proportion of women in this context is higher than compared to the global profile. This is indicated by the increased brightness of the list row. (see Figure 4) The same mechanism makes the countries Switzerland and Belgium visually more salient for the given context.

4 Related work

The rubber sheet [14] as well as the table lens approach[12] present the first instances of dynamically scaling list or table entries based on user interaction, thus introducing the focus & context principle for these forms of data presentation. However, scaling in this case only serves to make the contents visually accessible and size does not, as in our case, encode quantitative information.

The InfoZoom software (see e.g.[16]) uses dynamic scaling of horizontal list entries as indicator of relative proportions as well as miniaturized data plots to visualize quantitative data. However, designed as a database exploration tool, it aims at a diagrammatic representation of the data. Undisputably more powerful and elaborate than our approach form a data exploration perspective, we believe that our strategy of reducing complexity is more user–friendly for browsing and navigation purposes. Moreover, additional visual parameters indicating unusualness or temporal dynamics are not present, as in our prototype.

4.1 Work in progress

Currently, we are developing a prototype web feed reader application which utilizes the described principle through-

out the whole user interface. It is based on a minimal conceptual model, where microcontent is organized in "feeds", which are temporally ordered sets of items representing web feeds, but also dynamic collections defined by metadata values or actions. In this model, a facet is constituted by a collection of feeds. By scaling tags co–occurring with the selected term, neighborhoods of related tags can be browsed seamlessly. Additionally, the scaling of top tags in the initial view allows an easy drill–down from a general to related and more specific tags.

A second field of application is the dynamic, contextualized access to learning objects in the domain of architecture and design. In the context of the MACE project ((Metadata for Architectural Contents in Europe), we plan to combine additional forms of visualizations for specific metadata types (such as location, time, theoretical concepts, but also usage, competence and context metadata) based on the described interaction and visualization principles.

5   Discussion

In this paper, we presented a prototype to demonstrate a novel user interface component for facet browsers called "elastic lists". It aims at enriching current interfaces with additional visual cues about the relative weights of metadata values, as well as how that weight differs from the global metadata distribution. Following a focus & context tradition in information visualization, filtered–out items never disappear completely, but are collapsed to a minimal height in animated transitions.

We are currently investigating how our approach can be used in different domains and for larger and more diverse datasets. Additionally, first experiments show that for the measure of unusualness, more elaborate mechanisms than a mere ratio comparison are needed. Currently, outliers and erroneous values are sometimes over–emphasized in comparison to more interesting values with heigher weight. We investigate how smoothing techniques or the integration of a confidence measure could be used to dampen these effects.

6   Acknowledgements

References

[1] Microformats initiative, http://microformats.org.

[2] Simile project: Longwell, http://simile.mit.edu/longwell/.

[3] R. Beale. Information fragments for a pervasive world. *SIGDOC*, 05.

[4] C. Cattuto, V. Loreto, and L. Pietronero. Collaborative tagging and semiotic dynamics, May 2006.

[5] T. Erickson, D. N. Smith, W. A. Kellogg, M. Laff, J. T. Richards, and E. Bradner. Socially translucent systems: social proxies, persistent conversation, and the design of 'babble'. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 72–79, New York, NY, USA, 1999. ACM Press.

[6] S. Golder and B. A. Huberman. The structure of collaborative tagging systems, Aug 2005.

[7] M. Hildebrand, J. van Ossenbruggen, and L. Hardman. /facet: A browser for heterogeneous semantic web repositories. In I. F. Cruz, S. Decker, D. Allemang, C. Preist, D. Schwabe, P. Mika, M. Uschold, and L. Aroyo, editors, *International Semantic Web Conference*, volume 4273 of *Lecture Notes in Computer Science*, pages 272–285. Springer, 2006.

[8] D. Karger and M. Schraefel. The pathetic fallacy of rdf, http://swui.semanticweb.org/swui06/papers/karger /pathetic_fallacy.html, 2006.

[9] D. A. Leene. Microcontent is everywhere! defining microcontent. MicroLearning 2006, June 2006.

[10] J. Nielsen. Microcontent: How to write headlines, page titles, and subject lines, http://www.useit.com/alertbox/980906.html, Sept. 1998.

[11] E. Quintarelli. Folksonomies: power to the people. June 2005.

[12] R. Rao and S. K. Card. The table lens: Merging graphical and symbolic representations in an interactive focus context visualization for tabular information. In *Proc. ACM Conf. Human Factors in Computing Systems, CHI*. ACM, 1994.

[13] G. M. Sacco. Dynamic taxonomies: A model for large information bases. *IEEE Trans. Knowl. Data Eng.*, 12(2):468–479, 2000.

[14] M. Sarkar, S. S. Snibbe, O. J. Tversky, and S. P. Reiss. Stretching the rubber sheet: A metaphor for viewing large layouts on small screens. In *ACM Symposium on User Interface Software and Technology*, pages 81–91, 1993.

[15] R. Sinha. A cognitive analysis of tagging, 2005.

[16] M. Spenke. Visualization and interactive analysis of blood parameters with infozoom. *Artificial Intelligence in Medicine*, 22(2):159–172, 2001.

[17] E. R. Tufte. *Beautiful Evidence*. Graphis Pr, 2006.

[18] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *CHI '03: Proceedings of the conference on Human factors in computing systems*, pages 401–408. ACM Press, 2003.