Mashing-up mashups: from collaborative mapping to community innovation toolkits

Novak, Jasminko, University of Zurich, Institute of Informatics, Information Management Research Group, Binzmühlestr. 14, 8050 Zurich, Switzerland, novak@ifi.unizh.ch

Voigt, Ben, University of Zurich, Institute of Informatics, Information Management Research Group, Binzmühlestr. 14, 8050 Zurich, Switzerland, voigt@ifi.unizh.ch

Abstract

Mashups are web applications combining content and functionality from different online sources via publicly available interfaces (e.g. API, RSS). This allows end-users to create new websites that dynamically combine services of existing providers. The website Programmable Web lists almost 800 such applications. Still, little work has analyzed their structural properties, design dimensions and socio-technical implications. In this paper we contribute to addressing this gap by proposing and applying a specific conceptual framework for analyzing and evaluating mashups. We discuss the results and implications of an exploratory study and identify design recommendations for building successful mashups as socio-technical systems.

Keywords: mashups, collaborative knowledge creation, communities, user-driven innovation

1 INTRODUCTION

Mashups are web applications combining data and functionality from different online sources via publicly available interfaces (e.g. API, RSS) (Merrill, 2006). A variety of open access sources provides access to information such as geographic data, pictures, videos, news or taxonomies. A combination of lightweight application development technologies (JavaScript, XML, DOM), web services and an increasingly open access to data sources allows developers with relatively little technical skills to combine existing content and functionalities in new ways, quickly and easily. The website Programmable Web (http://www.programmableweb.com) lists almost 800 mashup applications (with a 300% growth rate in last six months). In spite of such a great momentum of the "mashup movement", little work has analysed their structural properties, design dimensions and socio-technical implications. In this paper we contribute to addressing this gap by proposing and applying a conceptual framework for analyzing and evaluating mashup designs, in order to identify design recommendations for successful systems.

2 CONCEPTUAL FRAMEWORK

In order to describe and analyze structural properties of mashups as socio-technical systems and evaluate whether they can live up to their promise as enabling tools for new levels of user participation and innovation (O'Reilly, 2005), we propose a conceptual framework based on four main design dimensions: application type, technology, social infrastructure and degree of openness (Table 1).

APPLICATION TYPE	**TECHNOLOGY**	**SOCIAL INFRASTRUCTURE**	**OPENNESS**
Application Class	Data Acquisition Method	Social Information Enrichment	Service Interface
Purpose	Interface Technology	Enrichment Type	
Scope of Use	Context Input		
Data Source	Customization		

Table 1. Conceptual framework for analyzing and comparing structural properties of mashups

The application class is an open value attribute which sorts mashups based on their basic functional application area (e.g. mapping, search, photo, travel). The purpose attribute describes the primary class of intended use (e.g. information access, information sharing, commercial transaction) while scope signifies its primary intended use context (private, local, regional or global). The technology attributes describe a mashup in terms of how it acquires base content, what interface technologies it uses, how is user context determined and what kind of customization possibilities it offers. The data acquisition method distinguishes between building base content by automatic acquisition from an existing online source and manual content creation by the users, or a combination of both. The interface attribute maps the use of technologies such as Ajax, Java, Flash, HTML etc. Context input describes whether the users need to define their context of interest manually (e.g. clicking on a map, entering a search term) or if this can be inferred automatically (e.g. current geographic location, interest profiles).

The social infrastructure considers the extent to which the mashup use and functionality depends on social aspects. It is described primarily with respect to the level of social information enrichment (algorithmic, user, community or none) and the enrichment type that reflects the richness of user contributions (base content, comments, geotagging, tagging, rating, profiles or none). Openness describes whether data and/or functionalities of a given mashup can be used by others: the service interface distinguishes web services, RSS, XML-feeds, platform-specific APIs, other and none.

3 STUDY METHOD

The conceptual framework has been applied in an exploratory study of a sample of successful mashup applications in order to gain preliminary insights into structural properties of existing mashup designs and map the range of successful solutions. A special challenge has been the question: does the framework allow us to identify best practices and design recommendations for successful mashups?

The mashup community website the Programmable Web has been taken as the source for the analysis sample. It is the most complete point of reference of a constantly growing pool of mashup projects (open user submissions, growing daily). The sample has been determined by success indicators combining Google Page Rank (Page et al., 1998) with user visits and user ratings from the Programmable Web (taken in June & August 2006). First, only mashups with a Google Page Rank of 6 or higher and a user rating of 4.0 or higher (on a scale 1-5) have been selected. This produced a set of 105 entries. Such a selection reflects successfulness based both on an "expert" audience judgment (user community of the Programmable Web interested in mashups) and by the overall web impact. The sample has then been narrowed down by taking mashups with a Google Page Rank of 7 or higher, pruning duplicates (different installations of the same mashup) and defunct sites. Subsequently, mashups with a lower Google Page Rank but representing thriving or well-established sites in specific niche user communities were added. The resulting sample included 40 mashups considered successful by the criteria discussed above. The sample entries were analyzed according to the presented framework. The interpretation of results is based on qualitative and exploratory data analysis.

4 STUDY RESULTS

4.1 General Patterns

In the analyzed sample, mapping applications account for more than a third of all mashups (37%), with search as the second next application type covering only half of that percentage (19%) and the rest between 3%-7% (Fig. 1). The main application purpose is information access (59%) followed by information creation (20%) and information sharing (11%). Such applications range from different kinds of geographically mapped information for individual users and communities (e.g. city crime maps, housing maps, job search meta-engines, photo sharing services, tourism and outdoor itineraries, localized news and city-transit information) to meta-search services for informal, user-created content (e.g. blogs, podcasts) to alternative interfaces for online shopping and product information. A few applications also deal with information organization (5%) and commercial transactions (5%) such as payment and event management or mobile auction support.

The primary scope of use is equally divided between global (44%) vs. local and regional use (43%). A minority of private applications (13%) provide information or services for use exclusively by an individual and not shared with others: mostly for commercial transactions and information organization (60%). Main data sources are Google Maps (50%), Flickr (20%) and Del.icio.us (15%). with a long-tail including Google Search, Amazon, Yahoo, Wikipedia, Technorati and PayPal (5-7%).

The technology dimension shows a clear dominance of automatic data acquisition from existing sources (62%) and of Ajax (JavaScript + XML + DOM + REST or SOAP) as interface technology of choice (61%). There is a strong minority of manual content creation (20%) and combined strategies (18%). The latter tend to fall into the mapping application area providing interesting insights into success factors of different strategies of base content acquisition (see next section). Ajax is used mostly by mapping applications (48%) while applications deploying basic HTML interfaces (21%) are found largely in the search area (43%). The use of Java in combination with Ajax turns out to be the technology of choice for applications supporting mobile access (75% of Java usage and 100% of mobile applications). Two-thirds of Flash applications are photo mashups. Context input is dominantly manual (90%) with only a few exceptions (mobile mapping applications, 10%). Customization is low (64% none) with some information personalization (19%) and interface customization (12%).

Social information enrichment is practiced by two-thirds of all mashups (68%) with roughly equal proportion of community-based (30%) and individual user enrichment (28%). Algorithmic enrichment by metadata induced implicitly form user actions or ratings accounts only for a minority of cases (10%). This varies with the data acquisition method: half of the sites using automatic data acquisition offer no social enrichment (48%) and only a few rely on community-based approaches (12%), compared to user-based (24%) or algorithmic methods (16%). On the contrary, all sites using manual data acquisition implement social information enrichment and tend to be community-driven (75%).

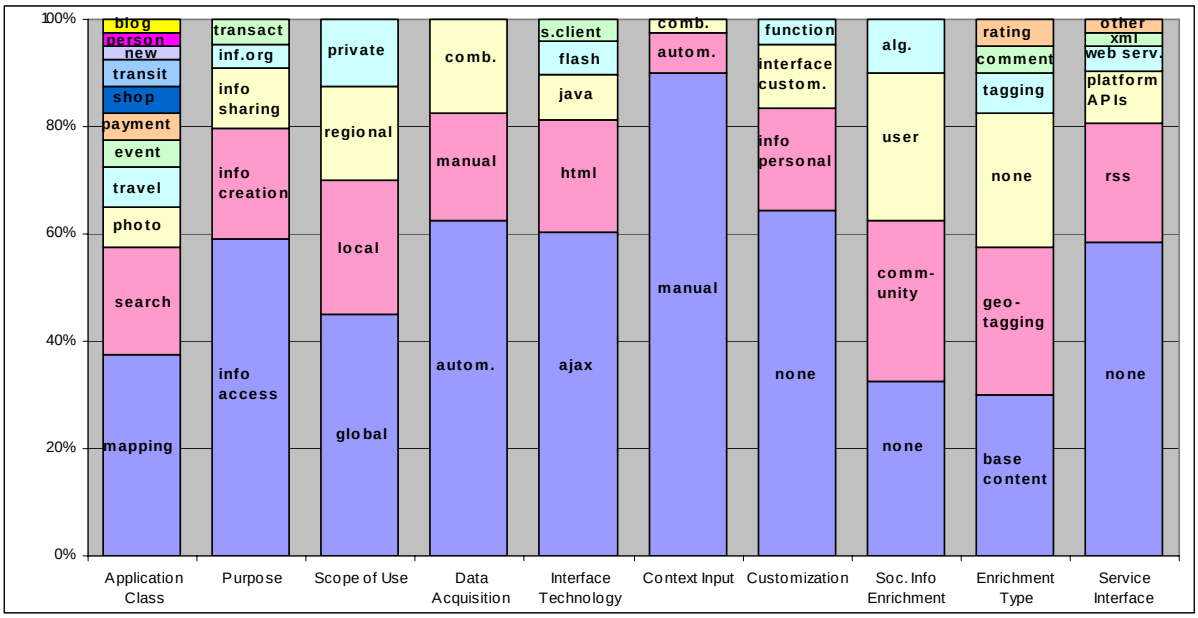

Fig. 1. Characterization of the overall "successful mashups" sample

Sites implementing a combined automatic and manual acquisition of base content are balanced between community-based and individual user enrichment (43% each). The most frequent enrichment types are the contributions to base content (29%, e.g. touristic descriptions, events or outdoor itineraries) and geotagging (28%). Geotagging is used across a spectrum of different applications (mapping, photo, shopping, travel, transit) albeit with strong dominance of the mapping area (55%). Non-geographic tagging, comments and rating appear only in a small minority of cases (blog meta-search sites). Finally, quite surprisingly, the openness of successful mashups is rather contained: the majority offers no access to their data or functionalities (59%). Only a few provide a web service interface (5%) or a platform-specific API (10%) while RSS is the main way of data syndication (22%).

The above analysis suggests an emerging dominant design pattern of succesful mashups across different application types: a combination of automatically acquired content base from existing sources with some degree of social information enrichment (largely geotagging). Customization and open access to mashup data and functionalities are low and don't seem to be critical success factors.

4.2 Mapping Applications

Mapping applications provide an interesting focus not only due to their dominance (with 37% by far the largest application class) but by the range of successful solutions differing from the overall dominant design. Mapping mashups typically combine different sources for data provision (e.g. cartographic services with classifieds, news, traffic or weather information), context information (e.g. geographic positioning) and user input and interaction (online or mobile) for access and enrichment of base-content with new knowledge. They all incorporate the principle of collaborative mapping, albeit in a range of different levels: from relating online content to physical locations (geotagging) to collaborative creation of primary, geographically relevant content to collaborative location-tracking. Geotagging is based on user input both in applications relying completely on user input for building the content base as well as in those using automatic acquisition from existing sources. In applications with manual data acquisition, user input goes beyond mere provision of physical location to creation of primary content, such as local information, user experience and advice (e.g. myoutdoors.net).

Mapping mashups based on automatic acquisition of base content from existing information sources (46,7%) are almost exclusively built for the purpose of providing geographically contextualised access to already available online information (85,7%) and rarely aim at new information creation (14,3%). In consequence, they generally provide no possibility of user information enrichment (71,2%). In applications based on manual data acquisition (33,3%) the content base is built from scratch, based

solely on user input. The purpose of such applications is largely information sharing (60%) and information creation (20%). The scope of use of the latter is exclusively local (66,7%) or regional (33,3%). Only a minority of mapping applications uses a combined strategy where the initial content base is automatically acquired from existing online sources and is then significantly extended by users' manual input of the same information type (20%). Furthermore, sites pursuing manual or combined content acquisition are largely community-based (80% and 66,7%, respectively). User information enrichment is primarily based on geo-tagging (40%) and base content addition (20%).

Due to its tight connection of the data source with a map-based presentation, the widespread use of Google Maps has introduced a common interaction and interface paradigm, becoming practically "overnight" a dominant design model. All mapping applications in the sample are based on Ajax and employ Google Maps as their application interface (100%). The input of user context, such as geographic location, is essentially manual. Rather then relying on sophisticated devices (e.g. GPS) the vast majority of applications requires users to explicitly pinpoint their location or region of interest on a map (73,3%). Only a few support automatic localization (26,7%). The majority focuses on online presentation and manual browsing of geographically contextualised content, rather then on personalised contextualisation by user location or profile.

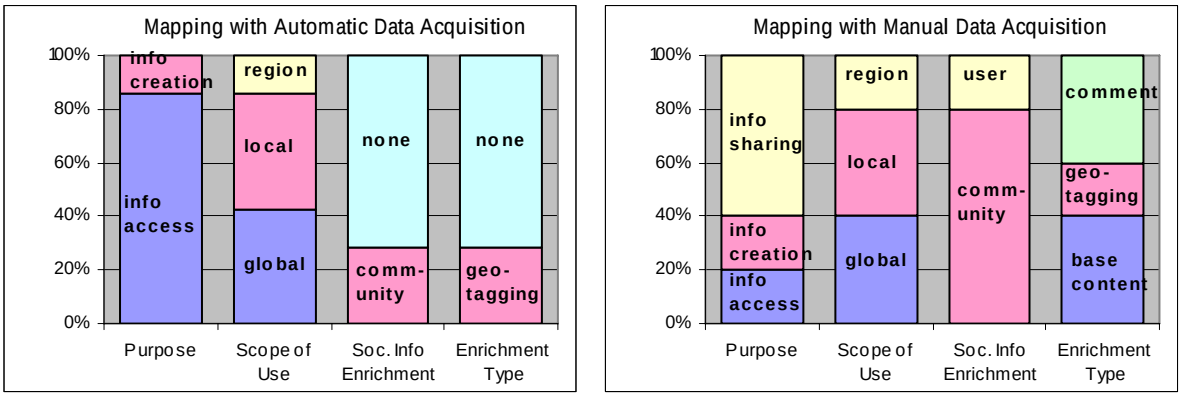

Fig. 2. Distinctive characteristics of mapping mashups

A distinctive characteristic of mapping mashups is their tendency to local scope of information use: 61% of sites offer geographically contextualised access to local (46,7%) or regional information (13,3%). The focus on local information use is also the defining shared characteristic of mashups with the purpose of information sharing (this purpose is found only in mapping applications). Locality is defined mostly in geographical terms or as a combination of locality in terms of a shared topic of interests with a geographical neighbourhood (e.g. bikely.com). This is observable also in some of the services with global scope of information use, such as frappr.com, which enables the creation of customized maps that allow community members to geographically visualize and locate each other. This serves both as a means for increasing awareness of an existing community as well as a means for making its existence visible to the outside and attracting new members. The map metaphor seems to respond to an intrinsic need and affinity of users to establish links between information and physical locations (53% of mapping sites employ manual or combined data acquisition). Spurred by the introduction of map-based interfaces that geographically contextualize shared information, communities originally based around shared interests regardless of physical location are re-discovering the physical dimension as an important element of the shared context.

A special class of mapping applications combines map-based information display with wiki-style content creation. Examples are collaboratively constructed tourism or travel sites (e.g. Wikimapia, Siam Soundtrek) or regional and local communities of interests (myoutdoors.net). Some are still struggling to reach critical mass while others have a well established content and user base. On one hand, this suggests an inherent advantage of designs that use an automatically acquired information base over those relying on user content creation. On the other hand, mashups targeting very specific information needs relevant to a certain user group (e.g. community) and related to a geographic area (e.g. city, region) have been successful in building the content and user base from scratch.

5 PRELIMINARY CONCLUSIONS

The presented exploratory study suggests the emergence of a general dominant design of successful mashups, where pre-produced information from available data sources (e.g. geographic maps, search engines, photo collections) is extended with manually created information by users in a collaborative manner. The degree of social enrichment varies between contributions by individuals, closed groups with moderation mechanisms and open communities. The extent of social enrichment varies greatly between application areas and is the highest for mapping applications. The latter also exhibit significant differences from the dominant design of the overall sample and provide insights into critical success factors of alternative strategies. They point to a niche design based on manual, community-based creation of the content base with high social information enrichment whose success is ensured by high specificity of information, closely related to needs of a specific user group *and* a geographically contained physical area.

Regarding openness, most mashups offer a limited degree of user customization and extendibility. Only a few provide APIs or data feeds that can be used by other services (33,4%). Openness is also intrinsically connected with the technological and social dimension. It has implications both for building or enhancing the content and user base as well as for personalization and customization capabilities. On one hand, information enrichment through geotagging is lagging behind the base cartographic data provided by existing sources just as is the provision of useful user-generated content (local information, experience and recommendations). Community sites focusing on manual creation and information sharing for very specific needs and local scope seem to fare better in this respect.

On the other hand, enabling the exchange of user created content and information enrichment between different mashups could provide the mapping community with a rich shared information pool needed for solving the cold-start problem and implementing effective personalization and recommendation services. However, the problem of semantic integration for mashup data interchange yet needs to be addressed. The availability of a common source of geographical information (Google Maps) and RSS as a de-facto standard for tagging-syndication could facilitate the exchange of geo-tagging information, but exchange of base content will require semantic mediation. Given the success of social bookmarking platforms (Hammond et al., 2005) it is unclear why such strategies have not spread for geo-tagged i.e. mapping content. Socially, an open question is what kind of collaboration models are conceivable between sites creating fresh content through manual user input and sites relying on automatic data acquisition from existing online sources.

This suggests an open challenge for mashup design: supporting information and service interchange (e.g. identity management) in mashup-networks and enabling users to build own applications based on existing mashup services. By providing APIs of their composite functionalities, mashups themselves could provide higher-level building blocks, further reducing technological complexity. This could facilitate end-user development and customization (Lieberman et al., 2006), turning mashups themselves into community "innovation toolkits" (von Hippel, 2005).

References

Merrill, D. (2006), Mashups: The new breed of Web app, IBM Web Architecture Technical Library, http://www-128.ibm.com/developerworks/library/x-mashups.html

Page L., Brin, S., Motwani, R., Winograd, T. (1998), The PageRank citation ranking: bringing order to the Web, Stanford Digital Libraries Working Paper.

Hammond, T., Hannay, T., Lund, B., Scott, J. (2005), Social Bookmarking Tools (I): A General Review, D-Lib Magazine, 11 (4)

O'Reilly, T. (2005). What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

Lieberman, H., Paterno, F., Wulf, V. (2006) End-User Development. Springer

von Hippel, E. (2005). Democratizing Innovation. MIT Press