Mashups: An Emerging Content Aggregation Web2.0 Paradigm

Ujjal Marjit   **Shibsankar Jana**

Abstract

The Web evolves rapidly. Day by day & WWW is currently experiencing a revolutionary growth due to its increasing participative community software applications. This paper highlights an emerging application development paradigm on the WWW, called mashup. Which uses latest generation of WWW tools and services enable web users to generate web applications that combine content from multiple sources, and provide them as unique services by adding values that suit their situational needs. This type of web applications is referred to as a Mashup. As blogs have enabled anyone to become a publisher, mashups stimulate web development by allowing anyone to combine existing data to develop single interactive web application. Current applications of mashups include tracking of events such as crime, hurricanes, earthquakes, meta-search integration of data and media feeds, interactive games, and as an organizer for web resources in a single interface. The implications of this emerging web integration and structuring paradigm remains yet to be explored fully. This paper describes mashups from a number of angles, highlighting current developments while providing sufficient illustrations to indicate its potential implications and their whole status in programmable Web.The role of Mashups on the web are presented, and existing & open emerging tools that assist in Mashup development are introduced.

Keywords: Web2.0, Mashup, API, RSS, Wiki, Blog ,Widget

1.   Introduction

In the changing context of wired environment, Organizations/Libraries require to accommodate the increasing needs of collaborative efforts, because coordination of their activities eventually determines their own success. The web 2.0 in itself emphasizes on user collaboration and participation, as well as using the web as a single universal platform to facilitate the organization's activities that enable collaboration at anytime from anywhere (OReilly 2005) [1]. There are lot of technologies viz; RSS, RIA, Wiki, folksonomy , Tagging, Mashup in Web 2.0.The ultimate Web 2.0 technology describe the efforts around content aggregation in a single web page commonly referred to as "Mashup". According to [2], a mashup comprises an application that "combines multiple sets of data streams into a unified user experience". Mashup approaches can be observed in many different fields, e.g. for enterprise information systems [3] or digital libraries [2], open digital journal (JUCS) etc. If we want to integrate online data from multiple sources to create a customize interactive web application without visit more than one sites then we can do it by using free, readily available tools & API that is why the hybrid online application introduces. People are using Mashups to mix the various types of data in a single web page. Newer Web 2.0 technologies are encouraging the mashups. Recent developments of technologies like wide deployment of XML web services,

Evolution of web browsers to enable greater extensibility (for example, Firefox extensions and Greasemonkey scripts), Widespread current interest in data remixing, Ajax and the availability of JavaScript-based widgets and micro-applications' Wider conceptualization of the Internet as a platform ("Web 2.0"), increased broadband access bring us closer to enabling users to recombine digital content and services. Moreover, mashups reflect deeper trends, even the deepest trends of human desire. As the quality, quantity, and diversity of information grow, users long for tools to access and manage this bewildering array of information. Many users will ultimately be satisfied by nothing less than an information environment that gives them seamless access to any digital content source, handles any content type, and applies any software service to this content. Mashups are web applications that combine information from a number of sources to provide a new user experience. This information isn't located at a single source; rather, it comes from numerous sources that publish information publicly through web services, RSS feeds, or other means. Traditionally, mashups involve combining

such information with a map. Mashups are a relatively new phenomenon on the Internet. The term is attributed to a similar concept in pop music where, in its most common form, the vocal tracks of one song were layered over the backing track of a completely different song. Such as Housing Maps (www.housingmaps.com), which combines housing listings from Craig's List (www.craigslist.org) with a map generated by Google Maps. The map is used to show locations where there are listings as well as the addresses and photos of available properties. Wikis play a role in mash-ups, as well (a

combination of content from various websites). For instance, Placeopedia1 combines data from Wikipedia and Google Maps. In future, it is possible that Wikis will primarily be used as modules in communications packages. Library 2.0 (Library +Web 2.0) is a mashup which includes wikis, streaming media, content aggregrators, instant messaging and social networks.

2.   Mashup is Different from Other Tools

2.1 Mashups Versus Traditional Aggregation (SOA Web1.0)

Mashups is a de-facto standards lightweight integration integrator whereas traditional integration based on tight integration which is risky and costly. Mashup uses the web as a platform which follows the new technologies like AJAX, RSS, JSON, REST, Micro formats, Atom, that is why it is a versatile and multipurpose content integrator. On the other hand traditional integrator which follows some common proprietary APIS, RPC, Message service. It is a very cumbersome and integration is big project. Traditional integration follows SOA Web1.0 technology but Mashups follows Web2.0 technology. Mashups is a client centric, web based adhoc aggregator whereas SOA is a server centric, well defined systems.SOA connects the systems whereas Mashups connect the usres.

2.2 Mashups Versus Portal

In terms of classification portal is an older technology and extension to traditional web server model, Mashup is a newer Web2.0 technology. Portal approaches aggregation by splitting role of web server into two phases- markup generation and aggregation of markup fragments whereas Mashup

adopts a more fundamental approach to content aggregation without regard to markup.The content aggregation of Portal takes place on the server but the content aggregation of Mashup can take place either on server or on the client.Portal aggregates presentation oriented markup fragments such as HTML,WML, VoiceXML etc. on the other hand mashup operates only on pure XML content also on presentation oriented content such as HTML.Read and Update models in portal are defined through a specific portlet APIs.but mashup CRUD operations are based on REST architectural principles but no formal APIs exists.

3. Architecture of the Mashup

Providers of APIs who make available components as building blocks for mashups are referred to as mashup enablers. The availability of simple and lightweight APIs has opened up a great number of possibilities for mashup assemblers.
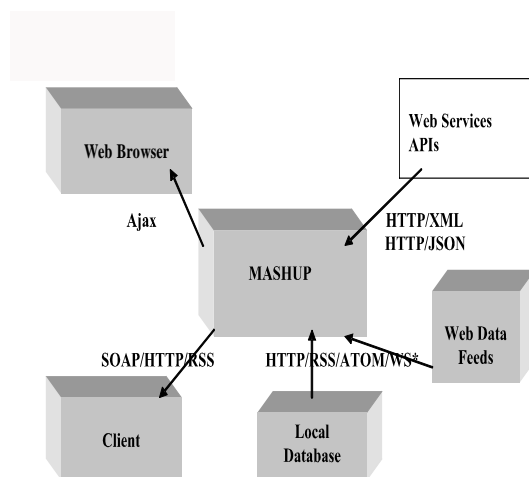

Figure-1: Architecture of the Mashup

Different type of data is remixed in mashup application using their API. Data for mashup is usually provided using an API from a third party,

but still it is the creativity of mashup creator that gives the combined provided services new value and semantic. RSS feeds, XML message which is retrieve in a SOAP envelop or local databases.
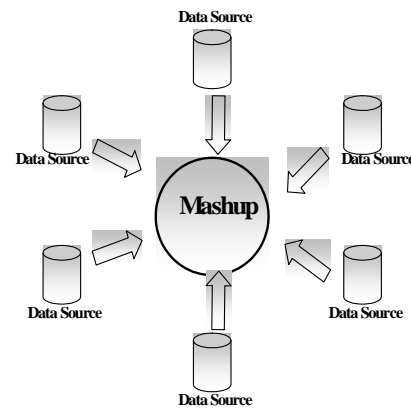

Figure-2: Data is remixed in a Mashup application

4. Typology of the Mashup

There are different types of typology of the mashup such as Presentation Mashups, Client-side software Masups, Client-side Data Masups, Server-side data Mashups, Server-side software mashups.

5. The Role of Mashups on the Web

One of the best resources on the web on Mashups is ProgrammableWeb 2, which contains links to Mashups and APIs as well as statistical data and Mash up-creating tutorials. ProgrammableWeb tracks more than 2800 Mashups and more than 740 APIs. However, compared to the daily increase rate of other Web 2.0 concepts like blogs, an average of 3.45 new Mashups a day as tracked by Programmable Web is disillusioning low. ProgrammableWeb also hosts the so-called "**Mashup Matrix",** a two-dimensional grid where both X- and Y-axis indicate an API. Thus each point in the grid is a representation of a tuple of
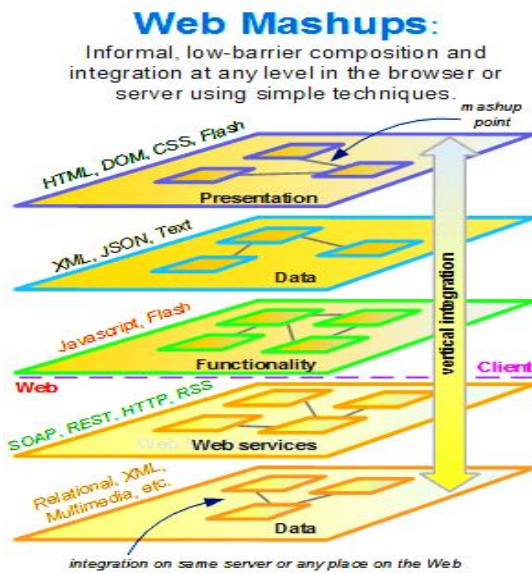
Figure-3: Web Mashups (Source: http://blogs.zdnet.com/ Hinchcliffe/?p=49)

APIs, which keeps a list of Mashups combining these two APIs. They illustrate the importance of Google Maps for the emerging of Mashups as a new web development paradigm

Data sources come from a range of Web content, including posted APIs, statistics, maps, RSS feeds, and advertisements. Mashup content is also sourced by 'screen scraping,' a process where, in the absence of an API, a computer program 'scrapes' a site for data, using code that crawls the site and collects the information in a format the programmer can use for his or her mashup. Many people are experimenting with mashups using Microsoft, Google, eBay, Amazon, Flickr, Serena, Facebook, and Yahoo APIs; companies often post their own API so that developers can utilize it in new mashups. The result is a value-added representation of data that makes it easier for a user to synthesize information.

John Musser's Programmable Web claims that over 50% of data mashups use Google Maps. Amazon has also started to allow access to its database through Amazon Web Services (AWS) API. Programmableweb [4] serves as an important resource for the categorisation of information and the analysis of an evolving collection of interesting and useful mash-ups.

The following chart shows the distribution of popular APIs that are used in this way:
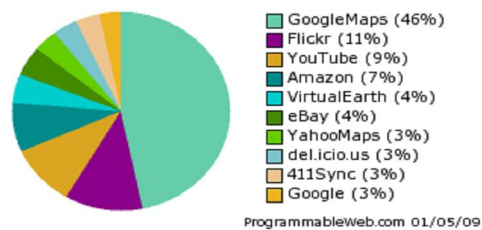
Top APIs for Mashups:

Figure-4: Top APIs for Mashups

Top Mashups Tags

Figure-5: Top Mashups Tags (Sources: programmableweb (http://www.programmableweb.com/mashups)

Mash up Matrix

The ProgrammableWeb Mashup Matrix (beta) lets you visualize the web mashup ecosystem. Hover over any cell to find mashups built by combining APIs on two axes. Click and drag to move around the matrix. Use the slider to zoom in and out. Click on the name of a mashup to see the details including screenshot, rating, and more. Enjoy.

6.   Pros & Cons of the Mashup

One of the main problems why the mass of web users shies of creating Mashups is that there is a

lack of robust development tools that make life for developers' easier. Not only is this fact keeping more users to create Mashups, but also the mixing of services by means of hacking JavaScript and server-side scripts seems to be cumbersome for the ones who dare to mash. Due to this fact, Mashups mostly incorporate only one external data source, while in theory multiple ones can be combined. Nevertheless will the remixing of different web services play an important role on the WWW in the near future and smart companies will soon offer the tools to make life for Mashup developers easier. Assembling Mashups is such a do-it-yourself process,representing in a strong way the technical and creative possibilities on the internet.However, due to the fact that various external sources are combined, availability becomes a crucial factor. So if one of the services used is out of operation, the whole Mashup can fail to work. For this reason it is often a temporary problem due to maintenance work at the company providing a part of the data but services could also shut down their system at all which could also lead to the death of the Mashup if no similar source can be found (otherwise at least reengineering efforts need to be accomplished). A Mashup developer should have in mind as well that programming interfaces could change.

7.　Existing Mash up Tools and their Comparisons

There exist a wide range of following Mashup building tools from both industry and academia:

- ♦ **MIT's Simile**
- ♦ **Dapper**
- ♦ **Yahoo's Pipes**
- ♦ **MS's Popfly**
- ♦ **CMU's Marmite**

- ♦ **Intel's Mashmaker**.

We will compare among the above-mentioned tool in terms of data retrieval, data cleaning and data integration. Simile the earliest system among all focuses mainly on retrieving the data from web pages using a DOM tree. Users can also tag sources with keywords that can be searched later. Dapper provides end-to-end system to build a mashup which is not available in Simile. However, users still have to do most of the work manually to define attributes and integrate data sources together. On the other hand, Dapper provides only one cleaning operation that enables users to extract a segment of text which is similar to Java's substring. Yahoo's Pipes, MS's Popfly, and CMU's Marmite [5] are similar structurally in terms of their approach. They rely on the widget paradigm where users select a widget, drop a widget onto a canvas, customize the widget, and specify how to connect widgets together. The Pipes system supports 43 numbers of widget where Popfly supports 300 numbers of widget. For example, Marmite and Popfly will suggest possible widgets that can be connected to existing ones on the canvas, while Pipes will rely on users to select the right widgets. MashMaker supports only web pages that are already extracted through Dapper. Intel's MashMaker took a different approach where its platform supports multiple levels of users. In MashMaker, expert users would do all the work in each area. Bungee Labs (www.bungeelabs.com), IBM's QED wiki (www.ibm.com), and Proto Software (www.protosw.com) are example Mashup tools for enterprise applications. These tools also use widgets to support most Mashup building functionality, but experts are required to use them because of their complexity. Desktop environment like OpenKapow (www.openkapow.com) allow users to 'sample' or 'cut' data from web pages to

be used later. However, both systems assume some level of expertise in programming in HTML and JavaScript. Apatar is the leading provider of open source software tools for the data integration market. With powerful Extract, Transform, and Load (ETL) capabilities, Apatar enables its users to easily link information between databases (such as MySQL, Microsoft SQL, Oracle), applications (Salesforce.com, SugarCRM), and the top Web 2.0 destinations (Flickr, Amazon S3). Apatar (www.apatar.com) provides support, training, and consulting services for its integration solutions. Headquartered in Western Massachusetts, Apatar operates a development center in Minsk, Belarus. Apatar is currently used by 3500 organizations and individuals worldwide.

8.   The Mashup Ecosystems

Figure-6: Mashup Ecosystem (Source: http://web2.wsj2.com)

9.   How might libraries use it?

Maps of local sites, historical sites, etc. • Additional information/combine info from more than once source in OPACs • Provide information specific to your locale .Libraries have skilled staff with professional expertise that can be leveraged to rise

to the challenge of Web 2.0, not only in collection and preservation, but also in usercentred services. They are also the guardians of a long tradition of a public service ethic which will increasingly be needed to deal with the privacy and legal issues raised by Web 2.0. Library staff should be encouraged to think and act pro-actively about how they can bring this to bear on the development of new, library and information service-based technologies. Librarians may introduce such type of technologies into the learning and academic workplace, driving the collaboration between academics, administrators and central information services.
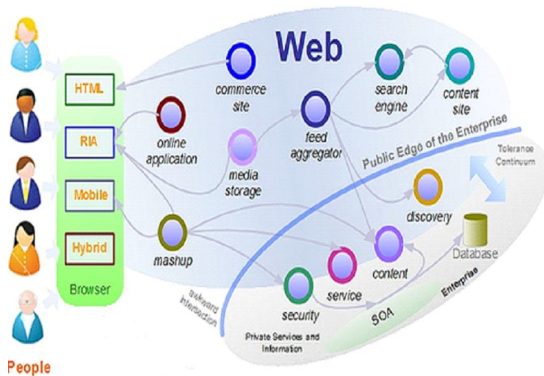
10. Mashups in the Future Web

The World Wide Web contains an immense amount of information, thus it is nowadays often thought of as a huge database. However, like for relational databases, a database management system (DBMS) is needed to combine data from different sources and give the information a new meaning. In above sections API driven Mashup building was introduced as a way of mixing up data from different Web sources just like combining data from different tables in a relational database, which provided a way of managing information stored in the database we call the World Wide Web. Building Mashups using API's require high programming skills though and so they are quite useless for a regular person, who wants to mix up data sources from all over the web. Another point is that most information on the Web is not accessible over an API, so only a small part of the WWW is remixable. In [7], the vision of gathering data for Mashups easier in the future is stated. "Web 3.0" is the keyword and its meaning is that "major web sites are going to be transformed into web services –

and will effectively expose their information to the world" [7]. Information, which is not accessible over an API, will be available using tools, which are already in use but still having a long way to go until they have reached a mature level.

Among these tools are online services called Dapper (http://www.dapper.net), Yahoo! Pipes, as well as desktop environments like Open Kapow (http://http://openkapow.com) and the already mentioned RSS managing tool. We have demonstrated the use of mashups to support collaborative content development. It could however be further extended to offer customized and personalized services to the academic community. Mashups could also be employed to facilitate further value-adds to digital journals and their communities of practice.

Future development of mashups can foster an interactive blended environment for organizations and institutions, providing multiple services for knowledge management. This includes the ability to visualize trails, data series, and events and the integration of multiple sources of contents in a variety of forms. Scholarly inputs can be incorporated to enable an identification of partners for collaboration, interesting papers in a specific area of research and to help gain in-depth insights on particular developments.

In future, meta-mashups (mashups of mashups) are expected to emerge as social contexts for community applications. Meta-mashups will integrate multiple (component) mashups together with other application such as Wikis. Meta-mashups will consolidate the input, output and processing of multiple sources of data into multi-layered architectures in the realization of complex applications. These mashups could also dynamically incorporate semantic modeling to contextualize or personalize applications. One of the biggest challenges in the development of mashups has been the lack of design tools to facilitate development by a larger number of end-users. Such tools are beginning to emerge, [7], but have not become widely adopted. These tools must become user friendly enough to allow non-computer experts to design useful applications.

11. Conclusion

This article discussed how libraries/Organizations can benefit from the new possibilities offered by Web 2.0 tools in content aggregation. Web 2.0 tools like Mashup can improve organizational and/or individual performance, but they also encounter a series of problems. Mshup is very much essential tool for creating the library web portal. We can develop multi featured mash-ups by incorporating more publications and bibliographic data from different journals, publicly available digital repositories such as CiteSeer, DBLP , Print Archives and by searching publications using other publicly available search APIs such as Google, Yahoo and other search engines. It is our belief that mash-ups will serve as an important paradigm shift for an electronic journal's expansion.Mashups provides "harmonization " of different software technologies.

References:

1. **Mashup (web application hybrid).** Available at http://en.wikipedia.org/w/index.php?title=Mashup _%28web_application_hybrid%29&oldid =98002063. (Accesed on 12/12/2008)

2. **Kulathuramaiyer, N. Mashups: Emerging Application Development Paradigm for a**

Digital Journal. Journal of Universal Computer Science. 2007.13(4), pp. 531—542

3. **7. Jhingran,** A. Enterprise Information Mashups: Integrating Information, Simply. 32nd International Conference on Very Large Data Bases, 3-4, VLDB Endowment, 2006.

4. **Programmableweb.** Available at http://www.progra mmableweb.com/Mashups. (Accesed on 06/01/2009)

5. **J. Wong and J.I. Hong.** Making Mashups with Marmite: Re-purposing Web Content through End-User Programming. ACM Conference on Human Factors in Computing Systems, CHI Letters, 2007. 9(1)

6. **Iskold, A.** Web 3.0 :When Web Sites Become Web Services. Available at http://www.readwriteweb.com/archives/Web_30_when_web_sites_become_web_services.php. (Accesed on 19/12/2008)

7. **Iskold, A.** Yahoo! Pipes and The Web As Database. Available at http://www.readwriteweb.com/archives/yahoo-pipesweb-database.php. (Accesed on 01/01/2009)

About Author

Mr. Ujjal Marjit

Mr. Sibsankar Jana, Assistant Librarian, University of Kalyani.

E-mail : sibs_jana@klyuniv.ac.in