8-5-2011

Role of mashups, social networking platforms and semantic in revolutionizing web integration: Key insights and enterprise implications

Krishna Chaitanya. T
Infosys Technologies Limited, KrishnaChaitanya_T@infosys.com

Jai Ganesh
Infosys Technologies Limited, jai_ganesh01@infosys.com

Role of mashups, social networking platforms and semantics in revolutionizing web integration: Key insights and enterprise implications

Krishna Chaitanya. T
Infosys Technologies Limited,
Hyderabad, India
KrishnaChaitanya_T@infosys.com

Jai Ganesh
Infosys Technologies Limited,
Bangalore, India
Jai_Ganesh01@infosys.com

ABSTRACT

The article examines the emergence and role of mashups in integrating data across the web, how social networking platforms penetrated the web by adding social experiences and how they helped in web integration, while solving some of the issues of mashups. We also discuss how semantic markups evolved in parallel, trying to bring in structured data and why social networking platforms need semantics for building the social graph of users. We take the case of Facebook's Open Graph protocol and discuss how semantic markups enhanced web integration by simplifying the process of creating mashups, giving more personalized experiences to the users. Finally, we present an integrated view of the role of mashups, social networking platforms and semantics from the angle of web integration and the insights that enterprises can learn from them.

Keywords

Web2.0, Mashups, Semantics, Social networking, Facebook, Open Graph protocol, RDFa, Microformats

INTRODUCTION

The World Wide Web has evolved from static textual documents, interconnected by hyperlinks, to a dynamic and programmable software platform (Mikkonen et al., 2010). During this evolution, several internet companies opened up their systems and exposed their data to the public via APIs (Application Programming Interfaces). An interesting consequence of this is web developers experimented with these APIs and created software applications called "Mashups", which combine content from multiple data sources on the web to create an integrated experience (Tuchinda et al., 2008). Along with mashups, social networking platforms and semantics evolved as some of the key building blocks of the next generation web, broadly referred to as Web 2.0. Web 2.0 is about the adoption of open technologies and architectural frameworks to facilitate participative computing (Ganesh, Padmanabhuni, 2007). Web 2.0 is shaping the way users work and interact with information on the web by shifting the focus to the user of the information. Mechanisms such as Mashups, Social Networking Platforms and Semantics are enablers for this. Web 2.0 has the potential to not only enable rich peer-to-peer interactions but also enable collaborative value creation across business partners. There are opportunities such as providing rich information on all the convergent services subscribed to by a consumer (including third party services) leveraging Web 2.0 standards which could be achieved through the use of Mashups based on content from multiple sources (exposed using APIS, RSS Feeds, Web Services etc.) to create new services (Ganesh, Padmanabhuni, 2007). The existing literature on Web 2.0 has examined issues such as online and social networking communities and their influence (Korica, Maurer and Schinagl, 2006; O'Marchu, Breslin and Decker, 2004, Kolbitsch and Hermann, 2006), benefits of social networks (Cross and Nohria, 2002; Garton and Haythornthwaite, 1997; Kautz and Selman, 1997), collaboration (McAfee, 2006) etc.

Mashups are one of the key building blocks of Web2.0, which are powered by technologies like AJAX (Garrett, 2005), data formats like XML, JSON (Crockford, 2001), architectures like REST (Fielding, 2000), DOM (Hegaret, 2002) parsing techniques etc. Though hundreds of APIs are emerging, mashups still existed only among the developer community, with lots of end user programming. Also, data portability was still something which could not be addressed completely. Social networking platforms are setting benchmarks in the way data can be exposed via their architecture and APIs. Their frameworks are so robust that thousands of clients are popping up to create useful apps. They are exhibiting the 'Lead by example' principle by building clean APIs first and then building their applications on top of their own APIs. The new Twitter client is a very good example for this. Apart from APIs, their social plugins and social sign on features are also penetrating the web. Semantics is a huge tree with each branch attracting a distinct community. Few people look at it as data

represented in RDF (Manola et al., 2004), OWL (Bao et al., 2009). For some it is about annotations in web pages with RDFa (Adida et al., 2008), Microformats (Allsopp, 2007) etc. Some are interested in semantic web services & APIs, some look at search, contextual technologies, AI etc., while business analysts look at value proposition offered by the applications to the end user. We are particularly interested in annotation of web pages, making use of the structural data for simplified web integration.

In this paper, we look at how social networking platforms and semantics are changing the evolution of mashups. We dig deeper to analyze the phenomenon by examining the evolution of the web till date. We first discuss how mashups emerged and played a crucial role in integrating data across the web and we list their issues. Next, we explain how social networking platforms penetrated the web by adding social experiences and how they helped in web integration, while solving some of the issues of mashups. In the next section we discuss how semantic markups evolved in parallel, trying to bring in structured data and why social networking platforms need semantics for building the taste graph of users. We take the case of Facebook's Open Graph protocol and explain how it is building the social graph of millions of users. Finally we explain the direction in which web integration is heading and the key technological areas which system developers and enterprises need to be cognizant of so that they can build systems which can scale conveniently.

HOW WEB INTEGRATION EVOLVED WITH MASHUPS

In Web 1.0 era, websites were static and data resided in silos. There was limited scope for sharing information with the outside world, with manual copy/paste or saving the web page. In Web 2.0 era, data of the silos could be leveraged, user-generated content became popular and collaboration became possible. Open APIs (Application Programming Interfaces which leverage modern web technologies for exposing data) started appearing which facilitated the creation of mashups thereby enabling information sharing. Traditionally, SOA (Service Oriented Architecture) was used to connect different systems and aggregate data. They used older programming languages and technologies that had more overhead, required more plumbing and infrastructure and hence their implementation was costly. E.g., Exchanging data over the wire in bulkier XML format needs more bandwidth and parsing time than JSON format, which is light weight and enjoys browser support for quick parsing. With enterprises opening up their APIs and evolution of newer technologies like AJAX, JSON, RSS, ATOM, REST etc., the integration became easier and programmers could easily mash up content using the web as the platform (Mikkonen et al., 2010). The focus shifted to small modular constituents which make the whole larger than the parts. There are distributed information packets which users can pull in and modify in new and innovative ways. The content sources could be exposed using APIS, Web Feeds, Web Services etc. There are no tight interconnections and there is facility for extension mechanisms enabling network participants to contribute and consume.

AJAX Start pages, widgets, mashup tools

During 2006, UI mashups took the popular form of widgets in AJAX start pages such as iGoogle, Page Flakes, Netvibes, ProtoPage etc. With simple HTML, CSS, JavaScript, widgets could be made pluggable into AJAX start pages. Mashups use content from multiple sources (exposed using APIs, Web Feeds, and Web Services etc.) to create new services. Content could be third party via a public interface or Data feeds such as RSS or Atom. Google, Ebay, Amazon have been publishing APIs that give access to their service. Availability of simple and lightweight API's has made mashups relatively easy to design. During 2007, mashups continued gathering attention in such a way that tech giants tried to build 'do-it-yourself' tools for enabling non developers to mash up content. Yahoo Pipes, Microsoft Popfly, Google Mashup Editor, Dapper etc., were some of the popular tools and these took care of data retrieval, data cleansing and data integration (Tuchinda et al., 2008). However, some of them could not last long and had to shut down for various business reasons, apart from lack of customization and scalability.

Problems in the mashups culture

Mashups brought up innumerable, creative possibilities which helped business leverage web technologies like never before and are still popular tools in the hands of developers. However, there are some inherent problems:

- In spite of open APIs coming into existence, web integration via mashups depends heavily on end user programming.
- Mashups allow shallow integration only at UI level and this helps mostly in visualizing data.
- The intelligent, user-generated content is still locked. i.e., data cannot be transferred beyond the point of its creation.

Yet, mashups continued to grow as they show value proposition to businesses. They had a newer dimension with the advent of modern social networking platforms. E.g., FaceConnector is an enterprise mashup which pulls Facebook profile and friend information into SalesForce CRM. Egypt protests mashup animates through latest tweets with #Egypt hash tag on Google

map. According to ProgrammableWeb.com, there are 21% mashups tagged as social, 12% search, 11% mapping etc., and the top APIs used for mashups show Google Maps API at 23%, Twitter at 10%, Facebook at 6% as on 27th Feb, 2011.

A snapshot of www.ProgrammableWeb.com as on Feb 22 2011.

THE EFFORT OF SOCIAL NETWORKING PLATFORMS

Social networking platforms have been scaling drastically, adding millions of users and their valuable data to their eco systems. While at one point, they existed as individual islands, they tried to talk to each other and exchange data with the evolution of oAuth. A number of social apps began to pop up, trying to mix and match the data between different social networks, giving rise to social mashups. While this improved the social networking platforms, it was still not sufficient for large scale web integration.

With respect to social networking platforms penetrating the web, Facebook is undoubtedly a juggernaut, with its rapidly growing user base, followed by twitter. The 2008 Facebook F8 conference announced 'Facebook Connect', which helped people to connect their Facebook identity with any external website and this was a major breakthrough. FB Connect's features such as Trusted Authentication, Real Identity, Friends Access and Dynamic Privacy helped in quick penetration into external websites arena. A couple of months later, twitter launched 'Sign in with Twitter' buttons, essentially with a similar concept of getting user base to an external website. These are different from mashups built out of open APIs by the mechanism that social networking platforms are now, in addition, authenticating/authorizing users, bringing their user base to a website and offering their services in the website itself- a step above simple mashing up of data.

During the 2010 F8 conference, Facebook announced 'Social plugins' and along the same time, Twitter came up with 'Twitter Anywhere' in their Chirp conference. These provide a way of embedding facebook/twitter widgets into websites, adding to the benefits of their login buttons which already had a positive impact. So after the age of AJAX start pages, social services poached to the doorsteps of web pages-a heavily accepted trend by developer community. The killer feature which dominated the web since 2010 Facebook F8 conference is "Like" buttons which spread across the web pages.

From their strategy of 'Like' buttons, it is very clear that Facebook doesn't just want to connect people, but it wants to connect people through the things they are interested across the web. This vision goes along the lines of Sir Tim Berners Lee's vision of the semantic web and Giant Global Graph. Re-imagining web as a graph of objects, in which people are interested, brings a new dimension to the evolution of the web and this is where Semantics contribute (Taylor et al., 2008). There were several parallel research projects by W3C and open source community on semantic web standards, but Facebook was the first one to put them into practical implementation on a large scale via their huge user base. They called their implementation of semantic web standards as "Open Graph Protocol (OGP)" and it completely changed the perception and notion of social networks. The ancestral hyperlinks of Web 1.0, which linked web pages, matured to 'Like' buttons of Web 2.0, which would design user's "social graph", a term coined by Mark Zuckerberg.

Thus, social networking platforms got seamlessly integrated into websites. They could solve some of the problems of mashups, listed at the end of section 2 above, by simplifying end user programming and providing means of exporting data

from the websites on which they were embedded. However, their APIs alone were not sufficient to share machine readable data and that's where semantics come into picture.

THE ROLE OF SEMANTICS IN WEB INTEGRATION:

The problem of designing a giant global graph of people and things boils down to, effectively describing the content of a web page, making it machine understandable. The applicability of semantics to social networking platforms could not be explained without a brief note on semantic web standards and their evolution.

Semantic web

As per Tim Berners Lee, the information about entities on a webpage and the relation between them can be represented using Resource Description Framework (RDF), which describes resources in Subject-Predicate-Object expressions called triples. Though RDF describes content completely, it is very complicated to build upon. To solve the complexity, open data format standard called Microformats was introduced. Instead of throwing away what works today, Microformats tried to solve problems using current usage patterns like XHTML (Mrissa et al., 2008). Though it received good adoption, its simplicity itself limited its extensibility. Along the same time, researchers came up with newer formats like eRDF, abMeta which enjoyed good support. W3C on the other side tried to extend RDF and proposed RDFa (RDF in attributes-Adida et al., 2008) in 2004, which makes it possible to embed RDF markup in XHTML and RDFa remained as the solution of choice for annotating markup. Though there was significant research in markup annotation, there were no big apps to use these. Recently, Yahoo and Google extended their support to RDFa and Microformats with Yahoo releasing 'Search Monkey' in 2008 and Google releasing 'Rich Snippets' in 2009. They were built on the concept of adjusting layout for specific type of search results. i.e., if the search query is an address, a map will be shown in results and if the query is a movie, reviews will be shown. In April 2010, Facebook announced their simple, RDFa based markup called 'Open Graph Protocol (OGP)' which allowed publishers to describe what object is there on the web page, helping in creating personalized experiences. During the same time Twitter announced 'Twitter annotations' in their Chirp conference, supporting RDFa, Microformats and OGP. It allows developers to add additional information (metadata) such as a string of text, location, URL etc., to a tweet, without exceeding 140 character counts. Though Rich snippets and other standards benefitted the web at large (Steiner, et al., 2010), Facebook's Open Graph protocol was the latest to join the list and it enjoyed a huge benefit owing to its huge user base and hence we shall analyze its impact.

The Open Graph Protocol (OGP) and why it is good for the web

The OGP enables any web page to become an object in Facebook social graph. In its current implementation, OGP is a simplified RDFa ontology implementation. OGP defines RDFa properties in the "og:" namespace which webmasters use in the <meta> tags, in the <head> section of their HTML files. There are four required properties for every webpage:

i. og:title- Title of the object as it should appear. E.g., 'Avatar'
ii. og:type- Type of the object. e.g., 'movie'
iii. og:image- Image URL which represents the object in social graph
iv. og:url- The URL which represents the object in social graph

e.g., For the movie 'Gladiator' on IMDB, the Open Graph Protocol's markup would look like:

```
<html xmlns:og="http://ogp.me/ns#">
    <head>
        <title>Gladiator (2000)</title>
        <meta property="og:title" content="Gladiator" />
        <meta property="og:type" content="movie" />
        <meta property="og:url" content="http://www.imdb.com/title/tt0172495/" />
        <meta property="og:image" content="http://ia.media-imdb.com/images/…../gladiator.jpg" />
        ...
    </head>
...
</html>
```

The OGP schema defines ~40 object types which include activities, businesses, groups, organizations, people, products & entertainment and websites. OGP also gives the ability to specify location, contact information, and attach audio/video data for describing the object in a more efficient way in the social graph. So, just by adding these four <meta> tags to a webpage, it can be converted to an object in FB's social graph.

Facebook Engineers designed the Open Graph Protocol and Facebook's Graph API so intelligently that placing a simple 'Like' button on a webpage can build the entire taste graph for users. There cannot be a simple way of spreading semantics for a huge platform like Facebook. The Graph API is a simple RESTful API uniformly representing objects (people, photos, events etc.) in the graph. When users click the 'Like' button in OGP enabled web pages, a graph API call is made to Facebook's servers, passing the User ID of the visitor and metadata about the page, which is extracted from OGP headers. This is a direct advantage for content publishers who ended up duplicating their content by setting up 'Facebook pages' for their marketing. Now they could enable their pages to behave exactly like FB pages, with very little development overhead.

Issues in OGP

- Since OGP is not having secondary attributes in its markup, it is not possible to differentiate between an object and its duplicate e.g., a movie and its remake and this started generating noise.
- There is strong criticism that Open Graph Protocol is not really 'open', since it was designed without collaboration with other communities. Also, it does not benefit anyone other than Facebook as of now, thereby helping them extend their data silos.
- Unlike Microformats, OGP cannot define multiple objects on a single web page, which is a setback in creating an ideal semantic web.

However, the fact that OGP increased the amount of semantic data on the web has to be acknowledged. Facebook would be trying to fix the loopholes very soon, since their protocol had huge success. Thus, semantics coupled with social networking platforms enabled in transferring machine readable data, which is critical for web integration.

One of the implications of this type of web integration is creating personalized experiences. Facebook released a feature called 'Instant personalization', initially limiting it to its partner sites- Pandora, Yelp and Docs.com. As per this feature, the partner sites chosen by Facebook can share the session created by FB's authentication system. The result is that, these partner sites will know user's public information on Facebook, without the need for an extra authentication layer on the site, by which they can create personalized experiences. Though instant personalization is in its infancy facing security and privacy concerns, it will definitely mature over a period of time. At a bird's view, is about using people's data for enhancing their experiences on the web and Facebook should surely be more open to continue their winning streak in the next decade.

INTEGRATED VIEW:

Web integration via the usage of social networks, semantics and mashups equally benefit enterprises, apart from general users of internet. Several new dimensions can be brought to the existing B2B, B2C, B2E, B2G spaces, which will enhance productivity, reduce costs, increase relevancy in several folds. To the external world, the efforts of enterprises in web integration will focus on taking the content to the users instead of taking users to the content, which is the custom in the present day enterprise. Within the enterprise, tasks become a lot easier if enterprises can accumulate structured data, which they can use for activities like quickly running a report, archiving, re-purposing, building widgets, make it more searchable, having mobile version of it etc.

Let us see how a future recruiting team of a B2B enterprise can leverage the technologies discussed in this paper. Their present applicant tracking systems work based on simple and often ineffective keyword searches through word documents and giving results which occasionally meet their requirements. Instead, if the enterprise uses modern web technologies, the catch would be close to perfect. If it embraces "hResume" Microformat, recruiters can automatically be alerted when more qualified applicants are available. Also, if it mashes applicants' online identity with Linked In, it can collect the recommendations of the applicant. Mashing up with slideshare can give the presentations given by the applicant, mashing up with specific tags on twitter can give the areas of expertise, mashing up with social bookmarking services for areas of interest and the possibilities are endless. In a B2C scenario, an upcoming ecommerce enterprise can benefit heavily with the usage of

data available via web integration. Techniques like 'Instant personalization', discussed in the paper above, are already in experimental stages, providing good results. Instead of adhoc advertisements on websites which often irritate customers, personalized suggestions can be made like what book to buy, which is the best high end computer system which matches user's needs and price range etc. Also, the demographic information of the customer can help in determining if the enterprise should target one specific audience or establishing various sites for different audience. The enterprise can get the information needed for this analysis from public social networks, by linking their authentication system with social logins and mashing up user data. Within an enterprise, adoption of Web 2.0 technologies and semantics can help in easing knowledge management process. Knowledge management is concerned with creating, maintaining and assessing the knowledge assets of an organization. It helps in establishing greater productivity, increased competency levels and competitiveness. With the global delivery model in place, many organizations are geographically dispersed and they use traditional document management systems for maintenance. These systems however have their own flaws e.g., they use keyword matching instead of intelligent semantic search, they retrieve information instead of answering queries, document exchange and customized view is not feasible etc. Instead, a semantic system in place can help in generating intelligence, apart from solving the above issues. A semantic wiki will be a more flexible solution in this scenario which can help in extracting, sharing content easily. These integration and intelligence aspects are some of the lessons enterprises can learn from the recent happenings in web integration, which are discussed in this paper.

CONCLUSION:

In this article, we have examined the direction in which web integration is heading. We have seen how mashups added value to the web by providing intelligent, user generated content. Then we have seen social networking as a platform for web integration, which moved a step above and helped in sharing user generated content to the outside world. We have also seen how semantics are helping in making the shared data more meaningful, taking Facebook and Open Graph Protocol as an example. For sure, we cannot call this as the conclusion for growth of the web and the big fight now is for gaining control over "user's data", which is becoming the currency on the web. Data portability on the web is one of the key areas of research gaining momentum, which is about users carrying their online data along with them, wherever they go. Projects such as DBPedia, Linking Open Data (LOD), vocabularies such as FOAF, SIOC etc. and Open Government initiatives have been showing good signs of a more open web in various researches. The web is more dynamic right now and businesses should definitely eye on these developments and be more flexible for change than before. While analyzing the trends in web integration in this paper, we tried to show how user data on the web is growing important day by day. We also examined how enterprises can tap in and utilize this rich data which is a win-win for both customers and enterprises. Mining intelligence from user generated content in social networks is our research in progress and going further we intent to extend our research to semantic wikis, blogs with HTML5 Microdata (Hickson, 2011) etc.

REFERENCES

1. Ganesh, J. and Padmanabhuni, S, (2007) Web 2.0: Conceptual Framework and Research Directions, Proceedings of the 13th Americas Conference on Information Systems (AMCIS 2007), Keystone, USA, Paper 332, 1-9.
2. Korica, P. Maurer, H and Schinagl, W. (2006), The Growing Importance of E-Communities on the Web. IADIS International Conference on Web Based Communities: 165-174
3. O'Marchu, I, Breslin, J, Decker, S. (2004), Online Social and Business Networking Communities. DERI Technical Report, Available at: http://www.deri.at/fileadmin/documents/DERI-TR-2004-08-11.pdf.
4. Kolbitsch, J and M, Hermann. (2006), The Transformation of the Web: How Emerging Communities Shape the Information We Consume. Journal of Universal Computer Science, 12(2): 187-213.
5. Cross, R., N. Nohria, et al. (2002). "Six Myths about Informal Networks and How To Overcome Them." Sloan Management Review 43(3): 67-76.
6. Garton, L., C. Haythornthwaite, et al. (1997). "Studying online social networks." Journal of Computer-Mediated Communication 3(1): 1-27.
7. Kautz, H., B. Selman, et al. (1997). "Referral Web: combining social networks and collaborative filtering." Communications of the ACM 40(3): 63-65.
8. McAfee, A. P. (2006). "Enterprise 2.0: The Dawn of Emergent Collaboration." MIT Sloan Management Review 47(3): 21-28.

9. Herman, I. (2010), "A year on the Semantic Web @ W3C", Semantic Web Meetup
10. Tuchinda, R., Szekely, P., Knoblock, C.A. (2008), "Building Mashups By Example" - IUI, ACM
11. Hausenblas, M. (2009), "Exploiting Linked Data For Building Web applications, IEEEIC
12. Mikkonen, T., Taivalsaari, A. (2010), "The mashware challenge: bridging the gap between web development and software engineering, FoSER, ACM
13. Ankolekar, A., Krötzsch, M., Tran, T., Vrandecic, D. (2007), "The Two Cultures", WWW
14. Str¨omberg, J. (2010), "Tools to Bring a Blog onto the Semantic Web", ACM USCCS
15. Steiner, T., Troncy, R., Hausenblas, M. (2010), "How Google is using linked data today and vision for tomorrow", Future Internet Assembly\
16. Häsel, M. (2011), "Opensocial: an enabler for social applications on the web", Communications of ACM, Jan 1 2011
17. Bizer, C., Heath, T., Ayers, D., Raimond, Y. (2007), "Interlinking Open Data on the Web"
18. Mika, P. (2004), "Social networks and the semantic web", WI
19. Mrissa, M., Al-Jabari, M., Thiran, P. (2008), "Using Microformats to Personalize Web Experience", ICWEWorkshop
20. Taylor, J., Evans, C., Segaran, T. (2008), "Creating Semantic Mashups", Web 2.0 Expo
21. Iwazume, M., Kaneiwa, K., Zettsu, K. (2008), "KC3 Browser: Semantic Mash-up and Link-free Browsing", WWW
22. Manola, F., Miller, E. (2004), "RDF Primer". W3C Recommendation: http://www.w3.org/TR/rdf-primer/
23. Bao, J., Kendall, E., McGuinness, D., Patel, P. (2009), "OWL 2 Web Ontology Language Quick Reference Guide": http://www.w3.org/TR/2009/REC-owl2-quick-reference-20091027/
24. Adida, B., Birbeck, M., McCarron, S., Pemberton, S. (2008), RDFa in XHTML: Syntax and Processing. W3C Recommendation, 14 Oct 2008. http://www.w3.org/TR/rdfa-syntax/
25. Allsopp, J (2007). "Microformats: Empowering Your Markup for Web 2.0"
26. Hickson, I. (2011), "HTML Microdata", W3C Working Draft 13 January: http://www.w3.org/TR/microdata/
27. Hegaret, P. (2002), The W3C Document Object Model (DOM)
28. Garrett, J. (2005), AJAX: a new approach to web applications
29. Crockford, D. (2001), JSON, http://json.org
30. Fielding, R. (2000), REST architecture (Doctoral dissertation)
31. Connect distributed data across the web - www.linkeddata.org
32. Programmable web, directory of APIs and mashups - http://programmableweb.com
33. Mark Zuckerberg., Facebook F8 conference, April 2010 - http://facebook.com/f8
34. The Open Graph Protocol: http://ogp.me/