Scalable Computing: Practice and Experience
Volume 9, Number 4, pp. 219–241. http://www.scpe.org
ISSN 1895-1767

WORLD WIDE WEB ON THE MOVE

DOMINIK FLEJTER*, TOMASZ KACZMAREK*, AND MAREK KOWALKIEWICZ†

Abstract. In this paper we provide an overview of key changes that happened on the Web in a few recent years. We start by analyzing changes occurring at the level of widely understood Web infrastructure (standards, computing, storage). Then, we focus on machine-oriented and user-centric trends in representation of information (both structured and unstructured). Next, we briefly discuss evolution of types of on-line functionalities and their access modes. Fourth component of the Web that we analyze is related to a few directions in actual usage of Web and its impact on social life. Final part of this paper is devoted to topics that span previous components such as driving forces, business models and privacy.

Key words: World Wide Web, evolution, Web infrastructure, Web data, Web resources, social Web, Web usage, business models

1. Introduction. World Wide Web is not only the biggest information repository in the history of humanity; it is also a dynamic and very quickly evolving universe consisting of people, businesses, applications, infrastructures and resources dynamically interacting with each other. This evolution results in an increasing complexity both of its individual components, and of the ecosystem formed by interplay of these elements. This paper provides an overview of evolving components of modern Web, mostly focusing on changes that happened in few last years.

1.1. Components of Evolving Web. In this paper we analyze evolution of the Web along four major areas (see Fig. 1.1). The first area (infrastructure), concerned with basic components and services that enable functioning of World Wide Web, is discussed in Section 2. It includes both hardware and software that enable different models of Web-based storage and computing, as well as basic standards (including file formats and communication protocols) that make Web-based communication possible. The next component, discussed in Section 3, concerns resources on the Web. It is mostly concerned with actual presence and representation methods of different kinds of Web content and data. The next area, covered in Section 4, focuses on functionalities available on the Web. It is concerned with the operations that users can perform on Web data and content, how they are accessible and how they can be combined. Finally, the forth component of proposed schematic view, described in Section 5, is related to usage scenarios of on-line systems. It focuses on which available functionalities people and businesses really use, and how important is their role in todays economical and social life.

These four components concern four distinct areas of contemporary World Wide Web. However, important interactions between them can be observed, as described in Section 6. They concern both driving forces of Web development, and issues that span multiple components, such as business models and privacy.

2. Infrastructure. At the very dawn of Internet its infrastructural level consisted mostly of wires and basic communication protocols and standards (such as DNS). Application protocols and data transfer formats were at their infancy, rather foreseen than fully developed. Over time it covered more complex components of Internet communication. Firstly, a number of standards of growing complexity such as HTML, JavaScript, CSS, XML, RDF and RSS appeared and became popular. Secondly, on-line documents storage became easier with no need to possess own servers: FTP and HTTP servers (including free options) became available to all Internet users and a number of alternative storage platforms (including blogs, on-line file sharing and social networking sites) became part of infrastructure. Thirdly, some basic Web computing platforms (e.g. Apache/MySQL/PHP, Python, RubyOnRails, ASP.NET) became omnipresent making deployment of Web applications easier.

Infrastructure is important as its availability at affordable rates or at no (direct) cost at all is one of the building blocks of all Internet activities. At some level of abstraction, we can perceive infrastructure as a large scale mechanism of demand accumulation to obtain economies of scale. As infrastructural components are required by everyone on the Web, keeping them shared by all makes technological and economical optimization possible. As the result of infrastructure availability, the entry barriers for new innovative business and social solutions acting on the top of them are lowered.

*Poznan University of Economics, Department of Information Systems, al. Niepodleglosci 10, 60-967 Poznan, Poland, {D.Flejter, T.Kaczmarek}@kie.ae.poznan.pl
†SAP Research CEC Brisbane, Level 12, 133 Mary Street, Brisbane QLD 4000, Australia, marek.kowalkiewicz@sap.com

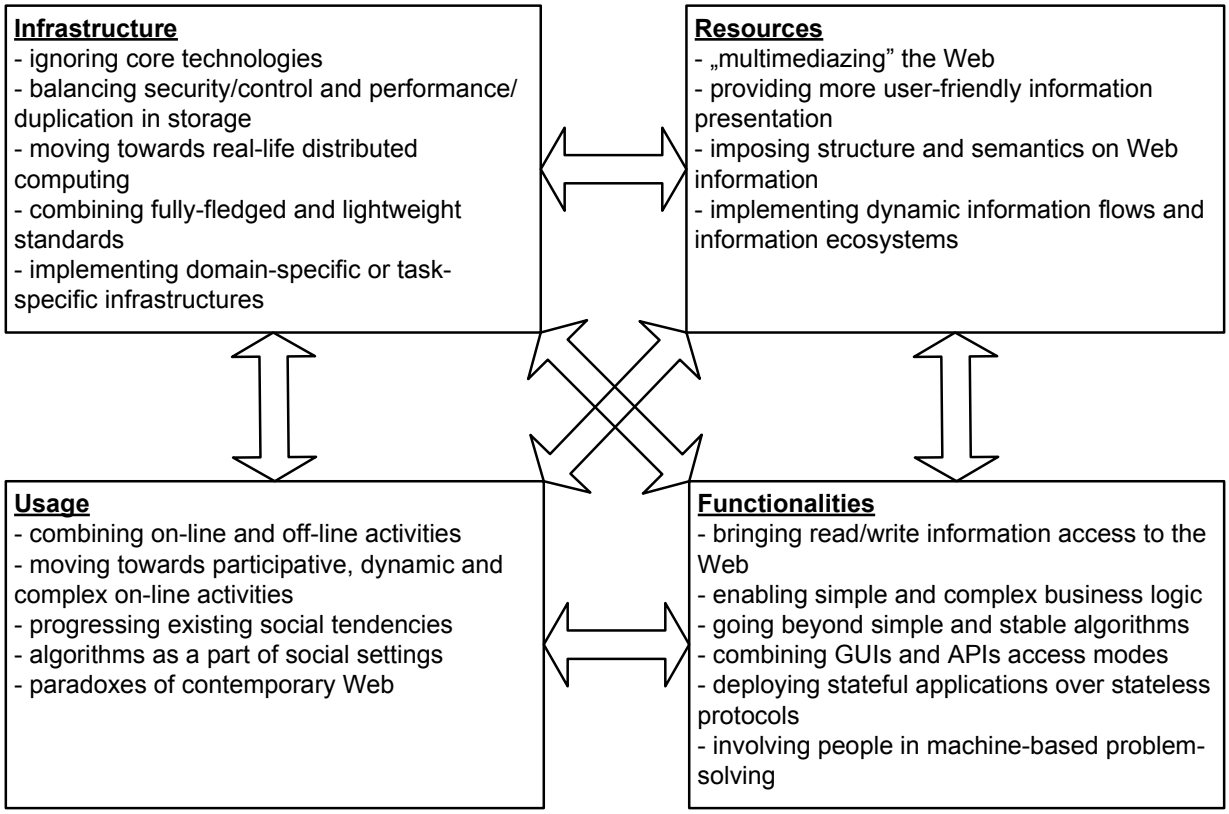

Fig. 1.1. *Four components of evolving Web*

2.1. Three Areas of Web Infrastructure. On-line infrastructure includes three main areas: standards, storage and computing. Standards propel all kind of communication and exchange on-line, thus they are the prerequisite for effective data flow, applications integration and business processes execution. In recent years we observe quick development of standards geared towards interoperability of data and distributed software.

Apart from standards developed or supported by standardization bodies (such as SOAP, RDF, OWL and OpenID), a number of formats and interoperability protocols (such as microformats, JSON or RESTful services) became *de facto* standards thanks to their wide adoption (see Table 2.1 for examples). They often solve the same problems as official standards in a less complete and flexible, but also simpler and easier to implement way. While the friction between competing standards causes confusion and bears new implementation challenges, it also results in better choice for developers and quicker maturing of new technologies.

Table 2.1
Fully-fledged standards and their lightweight counterparts

Area	**Fully-fledged standards**	**Lightweight (de facto) standards**
Remote calls	SOAP, CORBA, RMI	RESTful services, XML-RPC
Structural representation	XML	JSON, (X)HTML
Semantic representation	RDF, OWL, WSML	microformats in HTML
Federated identity	OpenID	e-mail as login
Metadata	Dublin Core	folksonomies
Portlets/Gadgets	JSR 286	Google Gadgets

One of the foundations of convergence of Web solutions, that we also perceive as a pseudo-standardization process, is the cultural tendency towards reusing best practices of other users and businesses. This results in similarities between business processes of many on-line businesses, multiple sites sharing similar information

organization schemas and even textual documents of specific types (e.g. advertisements or calls for papers) sharing their structure, formatting and layout features. It is worth noting that there are counter forces preventing total unification—they are driven by need for competition and differentiation of information both in terms of its content and processing capabilities, which are partially dependent on standards for information sharing. Thus we observe interesting process that throughout the years pushed the limit of standardization: first application protocols were agreed upon, later data representation formats were converging (this process is finalized currently), with final step in standardization of languages enabling flexible extensions to information representation formats, enabling both standardized processing and flexibility that enables value-added processing.

The second area of basic Web infrastructure consists of content and data storage facilities. It is shaped by two conflicting requirements, depicted in Figure 2.1. The former is to have maximal control over information location and its access rights—promoting storage centralization and forcing self-management (together with lack of affordable services to outsource storage). The latter requirement is to assure maximal performance (i. e. short access time from multiple locations, as well as storage scalability and persistence) and cost effectiveness—which is promoting distributed storage and outsourcing of the storage facilities.

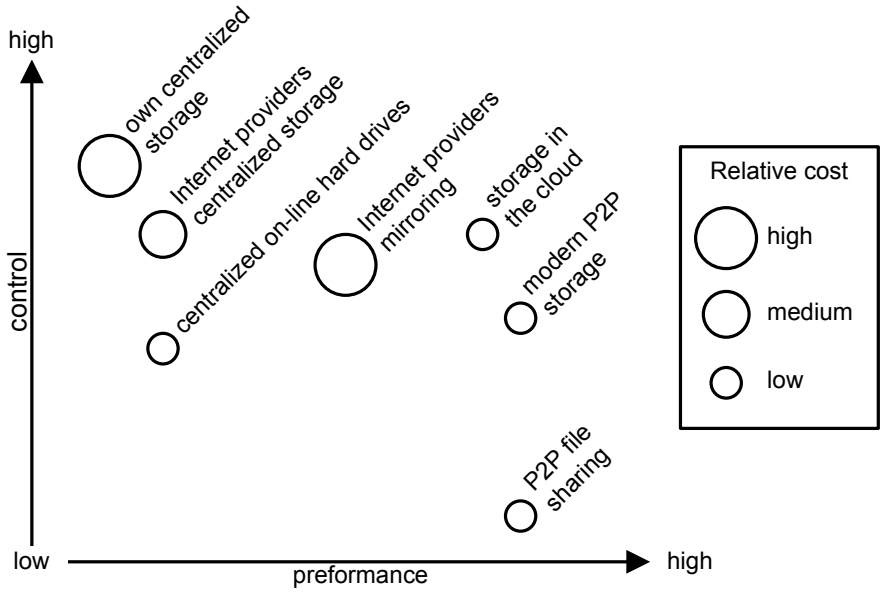

Fig. 2.1. *Control/performance tradeoff in storage solutions*

A few years ago the storage options were scarce: unless one was Yahoo! or Google, (s)he could only maintain own Web servers (typically more expensive and not necessarily more secure solution), or use individual servers made available by Internet providers (typically less expensive but somehow limited in functionality). In both cases, mirroring and using broadband connections were virtually the single options of increasing performance.

The first change we have witnesses was popularization of peer-to-peer (P2P) file-sharing applications. Peer-to-peer file storage proved scalable and assured rather good persistence of content (often against the will of its original creators or owners). However, in its pure form it also meant extreme lack of control over location and flow of information, making it absolutely inapplicable in business or personal information management scenarios. It is only after few adaptations, that P2P protocols found their way to the Web, making high-performance low-cost streaming of multimedia content more feasible (with BitTorrent being one of icons of this transformation). Today's distributed, cloud-based databases (such as Google's BigTable [7] or Amazon's SimpleDB[1] and Simple Storage Service[2]) learned lesson from both typical hosting and peer-to-peer systems, proposing what seems to be a good control-performance trade-off. Similarly as in case of hosting, the content is taken care of by a single

[1] `aws.amazon.com/simpledb/`.
[2] `http://aws.amazon.com/s3/`.

company. Similarly as in case of peer-to-peer systems information is distributed and replicated in multiple locations all over the world. However, in contrast to P2P networks, cloud-based storage is geographically stable and closed, thus does not suffer from high churn of nodes. Thanks to economies of scale, the proposed solutions are at least as affordable as hosting, with better performance, almost perfect scalability and usage-based cost calculation. In majority of cases distributed storage has higher uptime, even if spectacular failures happen (and may have high impact at least at the psychological level).[3] These failures encourage others (e.g. P2P storage Wuala[4]) to look for other solutions with a little bit more of a twist towards performance at expenses of control. Thus, cloud-based distributed storage surely is not the final answer to the centralized vs. distributed storage conflict.

The third area of Web infrastructure is related to on-line computing. Similarly as in case of storage, a few years ago this part of infrastructure was dominated by private or hosting-based servers using more or less standardized configurations (e.g. LAMP/WAMP[5] or Java-based technologies) to enable easy deployment of typical software solutions. Since then important changes occurred, leading to development of the rich computing environment that the Web is today. First revolutionary change is related to public accessibility of cloud computing platforms available form a number of companies including such huge players as Amazon (Amazon EC2[6]), Google (Google App Engine[7]) and Microsoft (Azure Services Platform[8]). These solutions, roughly classified as "platform as a service" (PaaS) solutions (e.g. Google App Engine, force.com) and "infrastructure as a service" (IaaS) solutions (e.g. Amazon EC2), make Web applications more scalable and available from all around the world. Moreover, as cloud computing platforms charge on per-usage basis, they are affordable for everyone and more economically reliable than previous solutions. Although spectacular failures of clouds generate a lot of fuzz, their uptime remains higher than for typical hosting solutions. In parallel, a shift towards virtualization enabled to build custom application stacks, and run them on multiple servers, or on cloud infrastructure (e.g. Amazon's EC2). Thus, today it is much easier to set up non-standard, scalable, high-performance servers, required by many specific Web-based services.[9]

2.2. Domain-Specific Infrastructures. Another rapid change at the edge of infrastructure is the development of domain-specific platforms that enable to build instances of specific applications with little effort. For examples ning[10] enables easy creation of social networking sites, Facebook Platform[11] enables development of applications using Facebook features and users base and Yahoo! BOSS[12] supports creation of custom search engines (and promises sharing revenue soon). A number of platforms for development of e-stores (including Yahoo! Store[13] and eBay Stores[14]) exist (the extreme example is Zlio.com[15] - in this case shop owner's activities are limited just to building a Web site and choosing product range; ordering, payment and logistics are supported by Zlio itself). Other examples include services such as TinyURL[16], Bit.ly[17] and purl[18] that aim at becoming another layer of standardized resources addressing on top of DNS. Another areas where some players aspire to become default infrastructure include enactment of complex information flows (Yahoo! Pipes[19] is the most renown example of such service), automated translation services (with tools such Google Translate[20] and Yahoo! BabelFish[21] competing with many smaller businesses), contextual ads (area strongly dominated

[3] See: http://www.readwriteweb.com/archives/google_failures_serious_time_t.php.
[4] http://www.wuala.com/.
[5] Linux/Windows + Apache + MySQL + PHP.
[6] Amazon Elastic Compute Cloud, http://aws.amazon.com/ec2/.
[7] http://code.google.com/appengine/.
[8] http://www.microsoft.com/azure/.
[9] Design of custom application stacks for virtual servers and cloud computing is simplified by services such as Elastic Server on Demand, http://elasticserver.com/.
[10] http://www.ning.com/.
[11] http://developers.facebook.com/.
[12] http://developer.yahoo.com/search/boss/.
[13] http://smallbusiness.yahoo.com/ecommerce/.
[14] http://stores.ebay.com/.
[15] http://www.zlio.com/
[16] http://www.tinyurl.com.
[17] http://bit.ly/.
[18] http://purl.org/.
[19] http://pipes.yahoo.com/.
[20] http://translate.google.com/.
[21] Originally developed for Altavista, now available at http://babelfish.yahoo.com/.

by Google AdWords[22]), on-line conference management (with EasyChair[23] being probably the dominant player and support for social network and content portability (with Gnip[24] being top commercial example, and SIOC community being the research leader [5]). It is also to be noted that a number of specific APIs were created with the objective of becoming standard infrastructure in specific applications areas. Examples include OpenCalais[25] from Reuters for natural language processing, Fire Eagle[26] for storing and manipulating location data, Mozilla Weave[27] - for storing and sharing data on browsing sessions, bookmarks etc. Finally, few infrastructure-like APIs focus on involving people into problem solving in multiple complex areas, such as information extraction, organization, integration and cleansing. This involvement takes multiple forms, including explicit (and paid for) people actions (as in case of Amazon Mechanical Turk service[28] or other forms of crowdsourcing different business activities including content creation, problem solving and even R&D [17]), and using analysis of behaviors of large groups of Internet users (for example in user reviews mining [18]).

3. Resources. The growth of size of resources available on-line has two faces: on one hand, we observe quick growth of quantity of content (i. e. unstructured information both in textual and multimedia form), on the other hand the Web is also the biggest repository of data (i. e. structured and semi-structured information). In both cases the changes are not only quantitative but also qualitative: the way data and content is made available on-line is evolving rapidly towards two (often opposed) objectives: one is representation better adjusted to needs of users and other is the form easily processable by machines. Example of these two tendencies are represented in Figure 3.1 and discussed in two following sections.

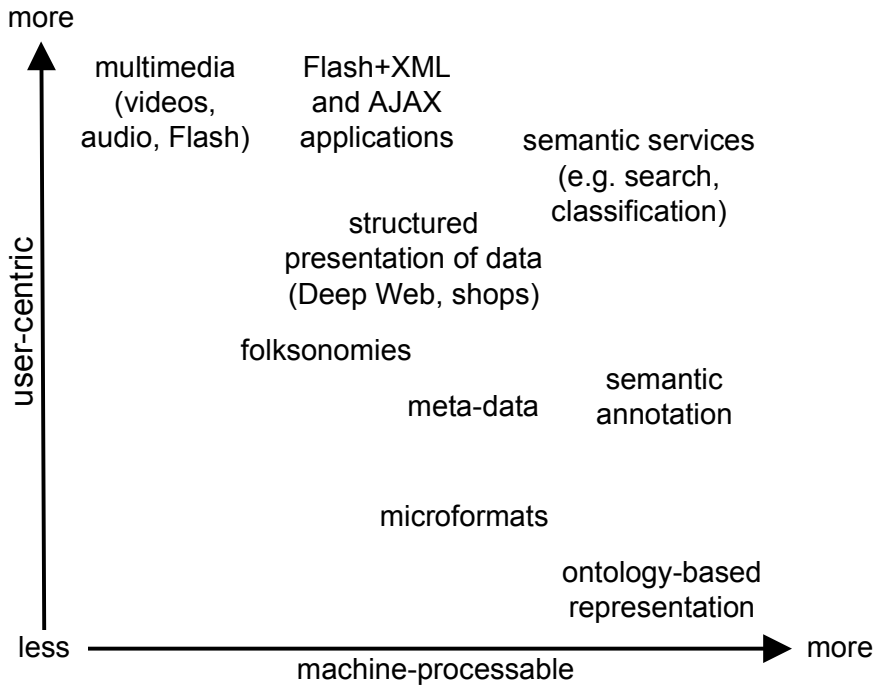

FIG. 3.1. *User-centric and machine-processability-oriented tendencies in functionalities*

3.1. User-centric tendencies. The tendency of making content and data more adapted to human users takes two main angles: on one side it affects the content and data themselves, on the other side it influences how the content is presented. Firstly, the online content in recent years has become more multimedia and visually

[22] http://adwords.google.com/.
[23] http://www.easychair.org.
[24] http://www.gnipcentral.com/.
[25] http://www.opencalais.com/.
[26] http://fireeagle.yahoo.net/.
[27] http://labs.mozilla.com/projects/weave/.
[28] https://www.mturk.com/.

appealing: thanks to wider broadband access, audio and video content become accessible to the vast part of Internet users. As a result more information previously provided in text form took much richer presentation: for example, growing part of software producers provides instructive videos apart from (or even instead of) text manuals, and more and more news are provided to users using podcasts. Secondly, thanks to such technologies as dHTML, AJAX[29], Adobe Flash and Silverlight, many on-line resources are not only multimedia, but also interactive and non-linear. These possibilities are for example widely used in different kinds of on-line training. It is to be noted that these changes often happen at the expense of accessibility, readability, and "skimability" of provided information, especially for people with special needs [8, 26].

The evolution of content presentation is mostly related to advance of dynamic user interfaces that try to mimic desktop software interaction paradigms (e.g. drag and drop or complex controls) and response times (e.g. by avoiding reload of the whole page on link click, using AJAX). Such rich user interfaces become a medium of its own, not easily separable from the content [24]. In extreme situations content is not a static, stable entity at all, and is recreated each time by a sequence of operations (happening both client and server-side) controlled by user actions, usage context (e.g. time, location of user) and external factors (e.g. other user's actions, real-life phenomena, random elements generated by algorithms). As an example - such dynamic content and its presentation is typical for real-time search engines (e.g. Twitter search) or highly personalized search facilities.

Similar tendency at the presentation layer is happening in case of some data-intensive Web sites. For example Flash or AJAX technologies are often used for interactive data selection or on-demand download of more details for already displayed data. In some cases, similar technologies are also used for visualization of data (not provided any more in textual form), for example as charts (in case of numerical data[30]) and simulations or models (e.g. in case of body colors in car industry).

Data and content presented using rich user interfaces are often called "dynamic". However, one more dimension of content and data dynamism should be also considered. Following the paradigm of collaboratively developed and maintained content, the growing amounts of information on-line are always-non-final, continuously evolving resources. This tendency touches even such traditionally stable entities as books (Wikibooks) or journal articles (scientific blogs). In parallel, mechanisms of partial control such as versioning and branching become popular. Similar dynamism can be observed in case of data; in this case quick changes result from tight connection to dynamic processes (e.g. in case of price lists, popularity of news articles or search result) or to measurement of dynamically evolving external conditions (e.g. sensor-based weather analysis). As a result, more and more content and data objects should be interpreted more as streams of new information (as in case of blogs, Twitter messages or sensor-based data sources) or of information updates (as in case of price lists and Wikipedia revisions), rather than stable entities.

3.2. Machine-processability tendencies. In parallel to the evolution of content and data format and presentation, we observe quick changes related to machine processability of information.

Firstly, a number of structure-centric, semantics-aware formats were proposed (as mentioned previously in the context of standard infrastructures). They may be used both to store metadata of on-line resources (e.g. title, categories, creator or tags of specific document), and to incorporate inline annotations into on-line documents (e.g. concerning specific named entities, numeric values or key phrases). Examples of used formats include family of XML technologies (XPath, XQuery, XSLT), RSS, RDF and OWL. Used meta-data and annotation schemes include microformats, Dublic Core, domain-specific ontologies such as FOAF and SIOC, MPEG-7 standard (for multimedia), and a number of non-standard annotation schemes proposed by different services.

Secondly, machine-processability of the content grow thanks to two complementary strategies to provide structure and semantics of data and content: bottom-up approach and top-down approach. In bottom-up approach the structure and semantics are imposed on content the by authors or Internet users by modification of underlying technology, or manual enrichment of the content. In most cases, this approach provides good quality structural and semantic information embedded directly in the content, typically keeping its human-readable character. While bottom-up approach is an important research topic and the number of sites that give some support to this approach is growing, the adoption of structured and semantic representation is still low, due

[29] While this term standard for asynchronous JavaScript and XML, it is also often used for asynchronous update of pages by using formats other than XML, such as JSON, XHTML or proprietary formats.

[30] See `http://www.tate.org.uk/netart/bvs/thedumpster.htm` for an interesting example for blogs visualization or Google Analytics motion charts

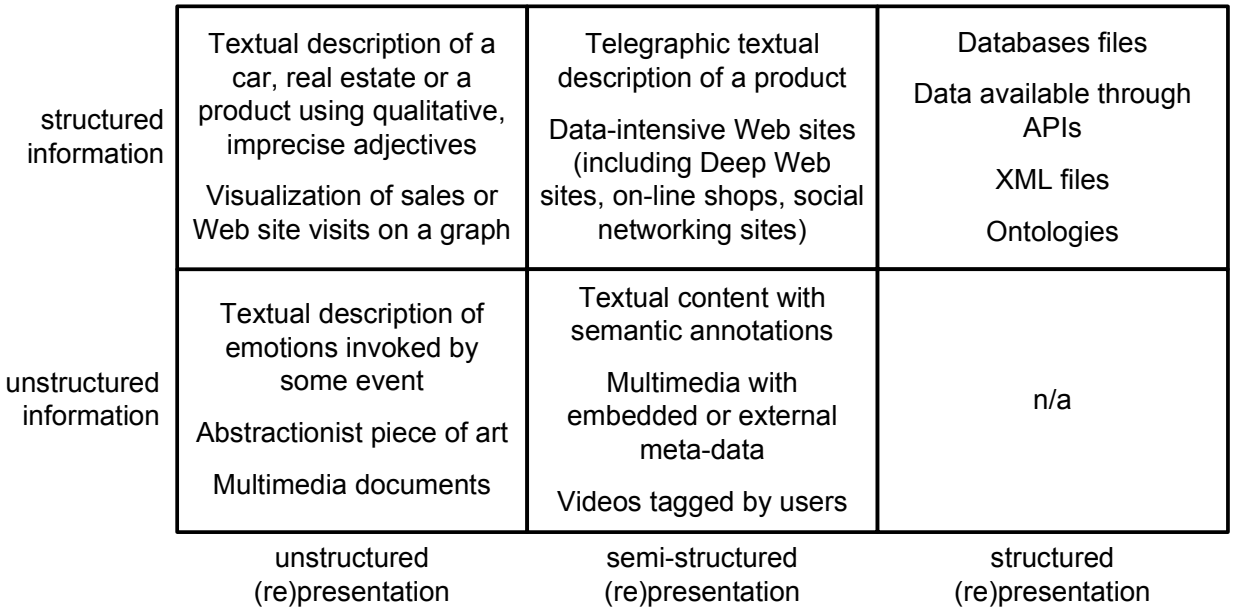

FIG. 3.2. *Levels of structure of information and presentation*

to weak incentives, relatively high costs and missing standardized vocabularies. All of these reasons propel the development of top-down approach, based on automated processing of Web content. In this approach the hints already present in the content are used together with external resources in order to structure and "semantify" information, without direct co-operation of individual Web sites. The output of automated top-down processing is more "digestible" (e.g. better structured, aggregated, organized or summarized) to users or machines than original content. This approach combines different techniques of Web content mining (such as classification and clustering, text summarization, information extraction, relations mining, ontology learning and population, opinion mining or multimedia content analysis), Web structure mining (such as measuring importance of Web sites, community discovery based on dense subgraphs of Web graph, Web site complexity measurement and Web pages categorization) and Web usage mining (such as discovery of customer clusters, analysis of products or documents popularity, improvement of collaborative filtering). [20, 23, 6] Search engines are an classical example of top-down approach—by using some general mining rules, they impose a specific ordering (by some measure of relevance to keywords and analysis of link graphs), specific structure (snippets reflecting contents of given page) and similarity-driven mechanisms (such as search for similar pages or clustering of results). First search engines used only Web content mining techniques. Then we observed Google's break-through PageRank algorithm using Web structure mining. Today major search engines use also to some extent behavioral analysis based on Web usage mining. On the other hand, a few smaller and ambitious players, such as Hakia[31], PowerSet[32] (recently acquired by Microsoft) and Evri[33], aim at enriching content not only with cited types of structure, but also with semantics. Some of today approaches to structuring Web content share properties of both bottom-up and top-down methods. Examples include social tagging and bookmark management sites such as del.icio.us[34], sites calculating other Web sites popularity based on votes such as social news site digg[35] or PostRank[36]—blog posts assessment service, different kinds of content annotation services such as SpinSpotter (allowing to annotate non-objective passages in newspaper articles)[37], and sites restructuring scripts or applications such as Dapper[38]. On one hand they are similar to top-down approach, because the structurization is happening outside of Web

[31] `http://www.hakia.com`.
[32] `http://www.powerset.com`.
[33] `http://www.evri.com/`.
[34] `http://delicious.com`
[35] `http://www.digg.com/`.
[36] `http://www.postrank.com`.
[37] See: `http://www.spinspotter.com/`.
[38] `http://www.dapper.net`.

site whose content is being structured, and because their approaches are general, often domain-independent, possibly large-scale and typically based on specialized algorithms. On the other hand, similarly to bottom-up approaches they are based on manual work rather than fully automated.

Thirdly, important contribution to machine processability comes from methods that enable both assessment of identity of multiple objects and measurement of their similarity are developed. Thus, both data and content objects are more and more connected and related to other entities. This area is strictly related to well-known research fields of schema mapping and matching of individual records or instances (which are a part of a number of information management tasks such as data integration, data cleansing and ontology merging). Moreover, quick progress in this areas influences both bottom-up and top-down solutions. Top-down solutions in schema mapping concern continuous improvements in methods of automated schema mapping. In recent years this area evolves towards holistic approaches, that enable mapping of multiple (often meaning: very large number of) schemas at once [14]. Related concept of "dataspaces" [13] seems to be implemented in real-life by Google Base, that gathered over 100K schemas and should allow large-scale schema mapping. A lot of top-down methods at the instance level were also proposed, using both more elaborate similarity measurement functions and better lexical resources. Significant body of research into ontology mapping and merging also fits to a large extent to this philosophy. On the other hand, ontologies are also the representation that promotes interconnection of multiple knowledge bases both at the class and instance level (linked data philosophy[39]). In bottom-up approach schema and ontology mappings and records equality are defined manually or by specific transformation software (varying from stand-alone procedural tools to declarative queries of rules executed by specific engines). Domain specific, dictionary-based records linkage is for example typical for shopping bots that help compare prices of the same products in different locations. Another helpful bottom-up tendency concerning instances is related to standardization of object properties formats (e.g. XBRL has been recently accepted by U.S. Securities and Exchange Commission as the required format for financial reports of public and mutual fund companies[40]) or to popularization of domain-specific identifiers (such as DOI[41] for electronic documents, or OpenID[42] for people). Bottom-up and top-down changes to the Web are happening simultaneously, and support one another. Even limited range of structure added to Web content may significantly lower the difficulty of top-down tasks. For example, usage of additional information encoded in user tags proved to be useful for Web content summarization [25], and potentially can have positive impact on performance of Web search [16]. Intuitively, when microformats are used, the task of information extraction (as well as tasks that depend on it, such as analysis of on-line social networks) should become much more feasible. Similarly, usage of tags may simplify the task of record linkage. On the other hand, top-down approach may significantly reduce costs of creation of semantic representation of content. It may even fully automate this process in some domains.

3.3. Content Flow and Content Ecosystems. One of characteristics of on-line data and content is their dynamic flow between a number of services. Originally posted to a single Web site (e.g. blog, shop price list or on-line database) or discussion list, the information may be reposted in a number of forms in other locations. Similarly, the changes to original content may be further propagated to a number of other locations. Complexity of such flows in case of blog posts is demonstrated by Figure 3.3.

The propagation of the content on the Web can be done by pop and push information flows. The former are initiated by the service that acquires a copy, and the latter is activated by information author or the service that the content is originally posted to. Examples of pop information flows include indexing by search engines or synchronization through RSS, examples of push flows include mirroring of content or submission of the same information to multiple Web sites.

In the same time the flows may be manual (fully performed by people), semi-automatic (requiring some setup activities but afterwards performed automatically) or fully automatic (requiring no user interaction at all). Examples of manual flows include quoting or copying content to other locations or forwarding it to friends. Examples of semi-automatic flows include mashups created with Yahoo! Pipes or YouTube videos embedded in a blog post. Typical examples of automatic flows are related to indexing and caching by search engines, or to usage of user comments on products for their automated qualitative assessment.

[39] http://linkeddata.org/.
[40] See: http://www.google.com/hostednews/ap/article/ALeqM5jTRoSiNGE5B07igsMWNH3ZOtbmAQD954M4800.
[41] http://www.doi.org.
[42] http://openid.net/.

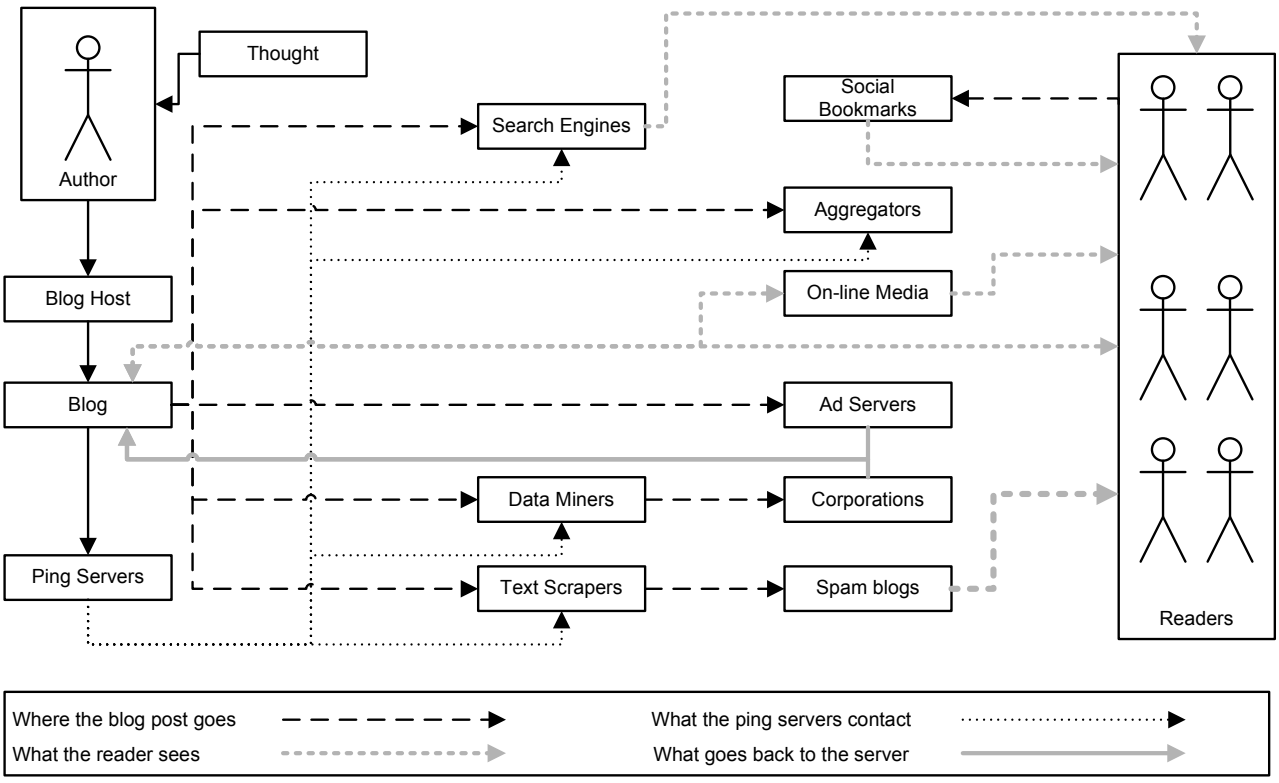

FIG. 3.3. *The life cycle of blog post as an example of complex flows (based on [27])*

Finally, the flows may preserve the identity of original content (e.g. in case of mirroring or embedding of content) or can do some transformations on content (e.g. adding semantics or hyperlinks, summarizing multiple user comments, or quoting a fragment of original text). For example automatic repost of e-mail group messages to Web-based archives or embedding of videos preserve the identity of original content. Machine translation services aim at providing the same content in different languages. Some NLP-based services (examples include eventSeer[43] in the area of calls for conference papers, or a plethora of services using OpenCalais) automatically add new links or meta-data. Finally, shopping bots, summarizing services (such as semantalyzR[44]) and other services using information extraction reuse only small portion of original content.

Different types of content flows are compared in Figure 3.4. It is also to be noted, that apart from flows that result in creation of new (instances of) content, information may be also accessible outside its original location via different types of querying services including meta-search engines, on-the-fly search tools (e.g. Twitter search[45]) or on-demand translations services (that do not store translated texts).

It will be interesting to see what will be the impact of content flows and ecosystems to information transparency. It seems that currently there is little care given by content aggregators and services that transform it to providing information about the original source (and additional metadata such as time of retrieval) of information and actual transformations performed. As the phenomenon matures and is wider applied in business scenarios we might observe new formats emerging to provide such metadata and forces driving to increase transparency of processing.

4. Functionalities. The area of functionalities is concerned with all kinds of operations that can be performed on on-line resources. This area has undergone major changes from the beginning of Internet era. It is visible even in the case of basic operations related to access to data or content. They concern both the "read" access to Web resources (that was possible from the beginning of WWW), and different types of "write"

[43] `http://www.eventseer.net/`.
[44] `http://semantalyzr.com/`.
[45] `http://search.twitter.com/`.

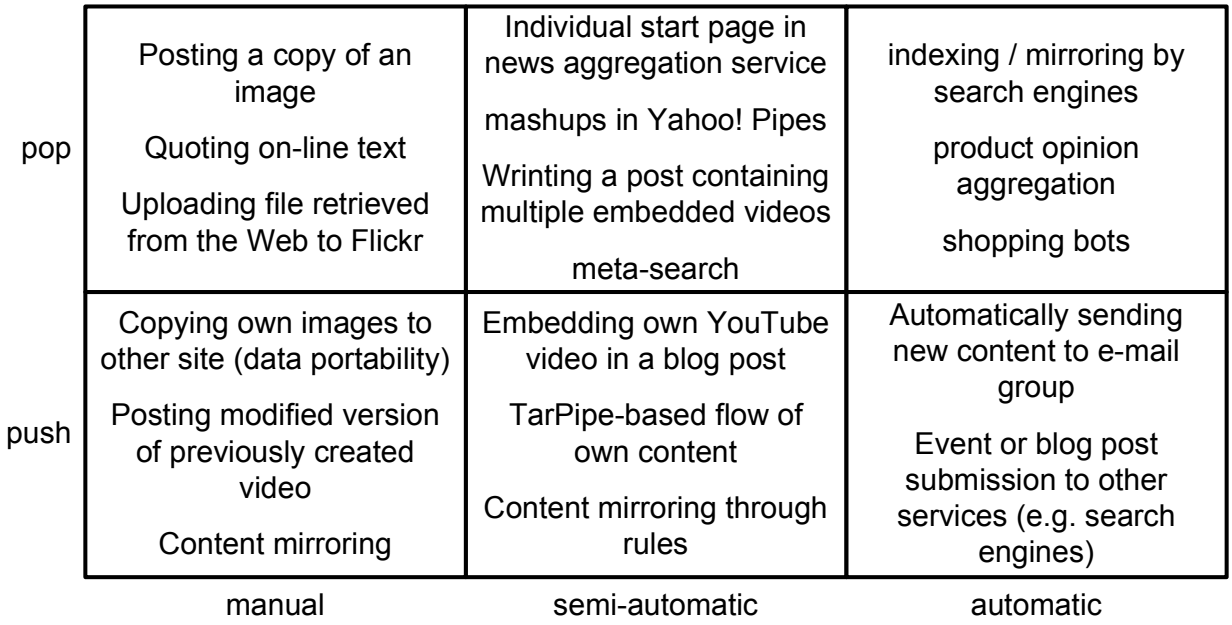

FIG. 3.4. *Examples of different types of information flows*

access (that were originally rarely supported on the Web, although foresaw by creators of the basic Internet infrastructure).

"Read" access to information has been transformed mainly thanks to already discussed bottom-up and top-down structurization and semantification of Web content and data. "Write" access to content and data includes relatively recently popularized features such as possibility of content editing (e.g. in case of wikis), creation of new content (also by duplication and edition of existing content), extension of existing content (e.g. by tagging resources, adding comments to articles or forum posts, or writing additional statements as in case of microblogging) and addition of new data, influencing aggregated data quality (e.g. voting in rankings, "digging" content, providing feedback on visited sights or hotels, or providing information on weather conditions[46]).

4.1. Business Logic Functionalities. Apart from such basic, storage-related functionalities, the on-line services implement different kind of business logic that focus on solving specific problems based on external or internal information. They include complex operations such as transformation, discovery, analysis, comparison, search and ranking of different types of information. The logic itself may have very different construction. It may be based on a stable algorithm (e.g. conversion of different measurement units, basic tax calculation), parameterized algorithm (e.g. tasks involving currency conversion, or tax calculation with changing tax rates or list of exempted products), algorithm applying user-provided rules (e.g. on-line content filtering based on preferences specialized by an user), machine learning algorithms (e.g. spam filtering functionalities), and interactive algorithms requiring participation of user or querying of external knowledge sources (e.g. search for ambiguous locations with search engines or map services). While majority of logic components provide exactly one and final resultset for specific input parameters, in some cases the logic may iteratively provide series of improved result sets (e.g. calculated with more iteration of optimization algorithms or constructed based on larger set of input data), based on a kind of subscription to results of on-demand calculation (like constant reordering of search results in some meta-search engines based on new data coming from multiple indexes).

While Web protocols are constructed as stateless, both stateless and stateful applications can be constructed on top of them. In case of stateless applications, activities (or invoked procedures) have no impact on results of future activities (or invocations) of the same user nor of other users. In stateful application current activities have impact on result of future activities with the same session (with the state stored temporarily), or also between sessions (with the state stored in a permanent way). The stored state can be itself meaningful to the

[46]For example in case of OtherWeather.com.

user (e.g. different aspects of manually typed user profile) or contain values that are solely machine-interpretable (e.g. vector representation of user interests based on keywords (s)he entered). Finally, the state may be attached to specific user (when login identification is used), to specific IP address, specific Web browser (using Cookies), some combination of the above, or may be shared by a number of users (e.g. the list of available tickets in on-line ticket sale service).

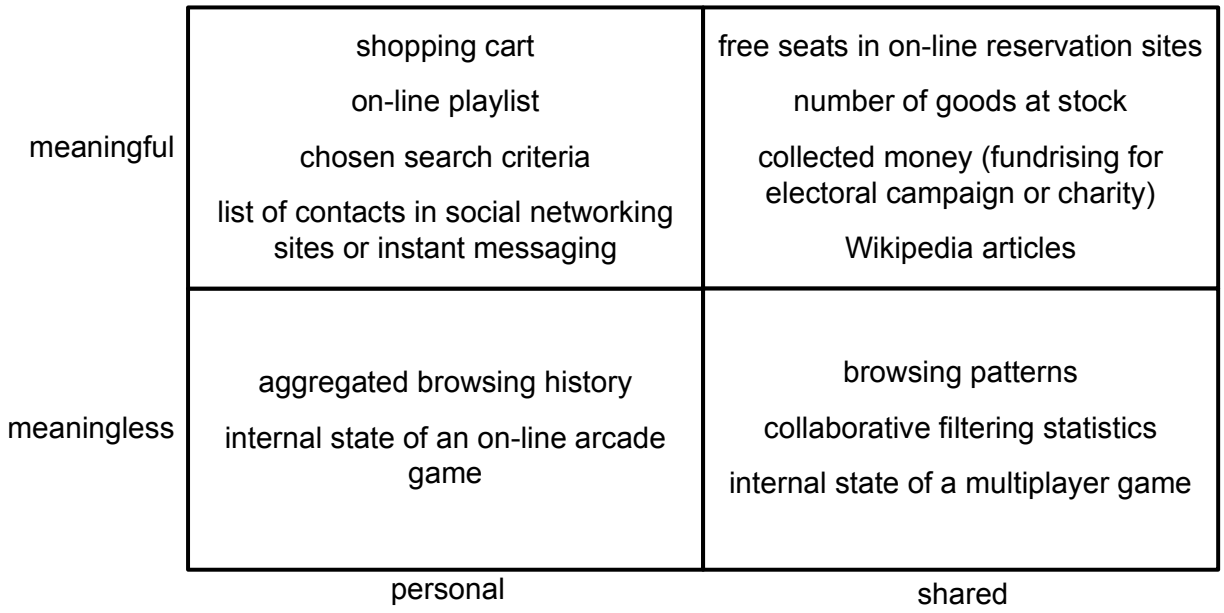

FIG. 4.1. *Examples of meaningful / meaningless and personal / shared state*

Figure 4.1 provides a number of examples for different types of state of business logic components.

4.2. Access Modes to On-line Functionalities. On today's Web, different functionalities are accessible in two basic modes: through Web-based GUIs and by different kinds of APIs. The first mode is focused on providing the access to features of on-line applications to human users. In this approach, Web operations are typically invoked by user entering specific pages, filling in forms or performing other HTML-based activities in Web browsers (such as clicking or dragging objects). However, some of these activities may use specific, non-Web technologies (such as Flash, Java or Sliverlight). While majority of logic in Web sites and Web applications is executed on server-side, more and more features are fully client-side. In many cases client logic is used as "a glue" combining functionalities provided by other server-side services (e.g. in case of mashups, widgets and embeddable JavaScript libraries such as Web analytics trackers). However, in some cases it may be accessible even purely in off-line mode (as in case of Google Gears[47]), blending the distinction between Web applications and desktop software. This blending goes even further with different types of business logic pluggable in user's browser with methods varying from lightweight (such as bookmarklets), through plugins using basically the same Web technologies but with greater access rights (e.g. FireFox plugins, Opera widgets, some Java and Flash applications), to fully integrated binary extensions such as Internet Explorer toolbars. At the extreme we can find desktop applications that embed Web browsers (e.g. for visualization or content access purposes) but have their logic hardcoded.

The second access mode, based on different type of APIs, is related to usage of on-line services by other software components. API types vary from complex and standardized (SOAP-based Web Services), through lightweight but mostly standardized (XML-RPC) to lightweight and mostly unstandardized (many REST-ful services with more or less stable and formalized response formats). It enables any applications to easily access and compose pieces of logic provided by multiple on-line services, as well as to access multiple types of on-line resources. Nowadays, this composition can be a part of client-side business logic of specific GUI-centric Web application (i.e. can be used internally by specific Web sites), it can be performed in a form of mashups

[47] `http://gears.google.com/`.

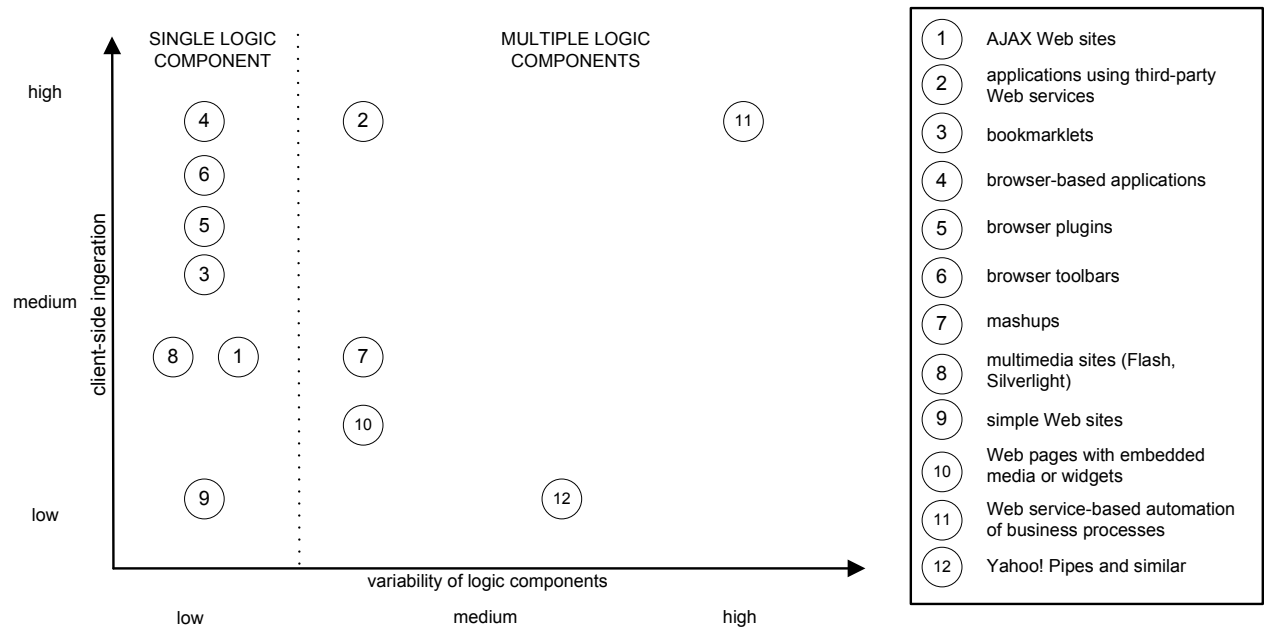

FIG. 4.2. *Different ways of implementing and exposing functionalities*

developed by programmers, or constructed using visual mashup construction tools (such as Yahoo Pipes). A lot of research also focuses on using semantics for composition based on technically underspecified business processes or objectives to be attained. It is to be noted that automated access to Web site content or functionalities may be also enabled if no API is provided, by using Web data extraction ("screen scraping") and navigation automation tools (such as Dapper, GreaseMonkey[48], WebVCR [2], iMacros[49]).

4.3. Involving People in Complex Functionalities. Typically when we think about business logic we mean automatically performed activities based on some pre-defined rules or algorithms. However, the open character of both Web and Web-based APIs makes the business logic potentially (co-)executed by people (individuals, businesses or groups of people). Human participation in composite logic may be synchronous (taking part at the moment of logic execution) or asynchronous (happening later). It may be also direct (with user actively taking decisions and actions) or indirect (with decisions being byproduct of other user activities, possibly aggregated over time by use of different machine learning methods). Few examples of direct and indirect, synchronous and asynchronous involvement of people in complex functionalities are gathered in Figure 4.3.

4.4. Examples of Typical On-line Functionalities. Typical on-line services focus on offering a few basic features. They include:

- information access and search—acquisition of information from Web sources,
- information management—management of document and media, including authoring, modifying, sharing, versioning, downloading,
- information transformation—transformation of one kind of information into different one,
- communication—spoken or written free-text exchange of information between people,
- collaboration and problem-solving—support for solving of complex problems by community,
- entertainment—individual or social hobbies, games etc.,
- self-development—education, training and spiritual development,
- business and transactions—acquisition and sale of goods and services on-line.

Apart from supported features two other dimensions may be used to classify Web sites. They are characteristics of the medium and properties of the application itself. Some of the most important properties of medium

[48] `http://www.greasespot.net/`.
[49] `http://imacros.net/`.

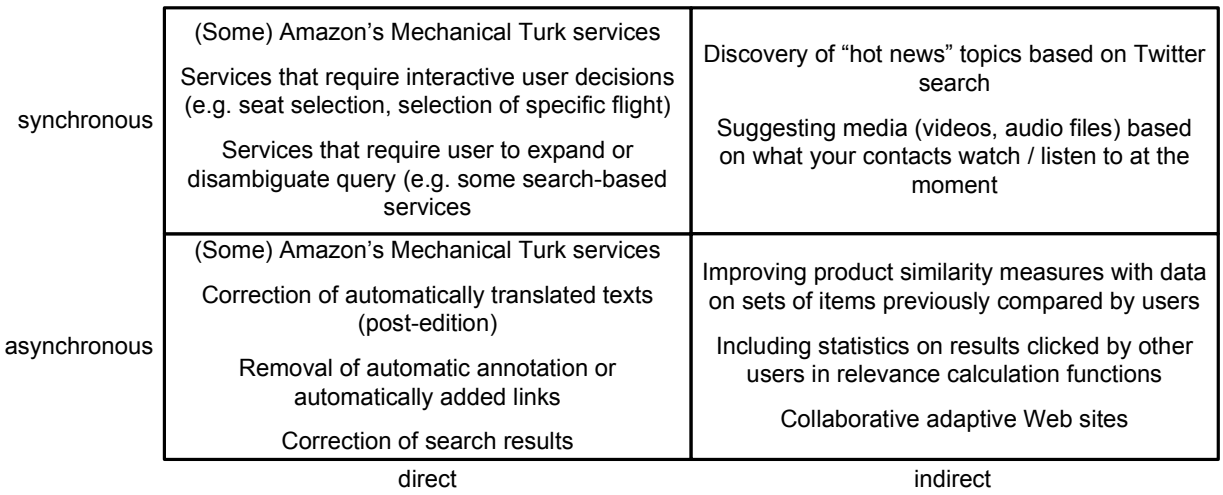

FIG. 4.3. *Examples of different ways of involving people in complex functionalities*

are: rhythm (synchronous / asynchronous), bandwidth consumption (low / high), format (text-based / voice) and permanence (persistent / ephemeral) [11].

Figure 4.4 compares a number of on-line services and classes of services with respect to the above dimensions and their support for aforementioned features.

4.5. New Web Paradigms Coming to Enterprises. The new functionalities and paradigms described above slowly pave their way to enterprises. On one hand, many of functionalities proposed by contemporary Web applications fit very well into quest for robust management of company knowledge. For example, wiki philosophy may be very useful in documentation creation and maintenance and may act as a supportive tool in project management; enterprise blogs can be an useful method of communication with both internal and external stakeholders (employees, shareholders, partners, customers, suppliers, potential customers). Finally, tagging (with unrestricted or partially restricted vocabularies) may be a more flexible alternative to other approaches of enterprise documents organization (such as classification and full-text indexing). At the same time, adoption of these "Enterprise 2.0" solutions is shaped by a structural conflict between openness and flexibility, typical for many modern Web-based systems, and control or rigid procedures, being landmark of contemporary enterprises.

On the other hand, we observe adoption of the paradigms related to aggregation of logic and information from multiple sources in the business scenarios. Over time more and more companies monitor and integrate information about company reputation, competitors actions and market changes from Web locations. While majority of businesses gather this information mostly for PR, marketing and strategic or tactical-level planning activities, the number of businesses using numerous integrated data sources in operational activities and the strategy of "competing on analytics" [10] is continuously growing. As more and more forms of inter-company collaboration is mediated by IT solutions, the flexible composition of logic from multiple providers (taking form of mashups, enterprise mashups, individual pipes or less loosely-coupled IT solutions) is progressing. This tendency starts to be supported by growing openness and service-orientation of major enterprise solution players (including SAP, Oracle and Microsoft). Finally, the ideas of simple, adaptive workflows combining automated activities with user involvement are becoming mainstream of research and are supposed to find their way to enterprises in closest future.

Despite this developments and buzz generated by Enterprise 2.0 solutions, majority of medium and large companies still operate multiple unintegrated or poorly integrated solutions (even if coming from the same provider) even internally. Moreover, many legacy IT software remain not well suited for or very restrictive about integration with external logic components (see for example [1]).

5. Usage. The area that recently changed the most from the point of view of people is the usage layer of the Web. It is concerned with what features of on-line applications are actually used and how. As it is an area of complex interaction between multiple systems and large number of users with very various background and

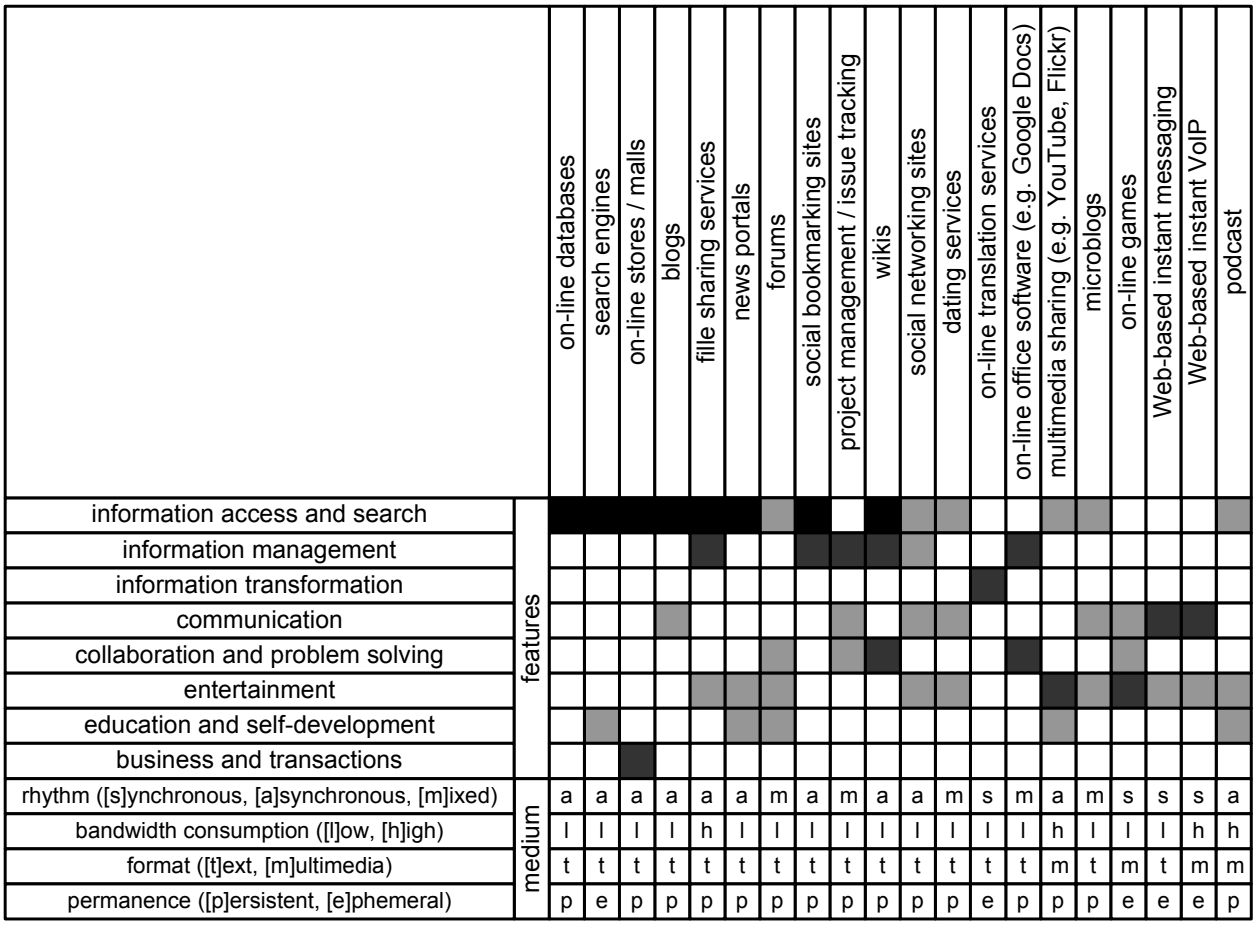

FIG. 4.4. *Examples of typical services on contemporary Web*

objectives, the "macro" impact of individual applications and their features at the social and economical level may be very hard to derive from their micro properties [15].

5.1. General Directions in Web Usage Evolution. Seen from somehow bigger distance, the Web evolves into a number of general directions. Typically this evolution means that new areas, that existed previously in an embryonic form, become mainstream of on-line businesses. At the same time, many of previously mainstream usage patterns still remain popular in specific types of services and groups of users. These general directions are:

1. Growing level of user engagement in on-line activities. Activities evolve from passive (e.g. browsing information) to active (involving participation in content creation). We identified five levels of engagement in content creation: a) no participation at all, b) unconscious participation (when patterns of users behavior are used in automatic content creation, as is in case of collaborative filtering or adaptive Web sites), c) participation in simple individual activities (e.g. tagging or rating URLs, products or blog posts), d) creative individual activities (e.g. writing blog posts or comments), and e) creative social activities (such as synchronous or asynchronous creation or management of long text documents, ontologies or databases).
2. Moving from individual to social activities. Until recently, the majority of computer-based activities were "single player". Today, a lot of them can be also done in a collaborative manner. For example, we switch from individual bookmarking, playlist management, searching, and problem solving to collaborating while performing these activities.
3. Moving from one-time to continuous and incremental activities. For example book and article writing

as well as film-making were one-time activities (after being finished the result did not change). Today even books (cf. Wikibooks) are editable in a wiki way and easy to comment on. At the same time, it became cheap to publish new versions of any digital content including multimedia and research papers. Moreover, editable and reusable content allows both the same user and other people to create new, improved or mashuped-up content. It is also a general tendency to create algorithms that first approximate results and then utilize user feedback on results (acquired through both implicit and explicit feedback cycle) for continuous result improvement. This last tendency might be due to strive to solve problems that are not tractable using traditional approach.

4. Moving from asynchronous to mixture of synchronous and asynchronous interaction mode. Majority of interaction on the Web used to be mediated by some content and performed in a asynchronous way. Recently we have witnessed, the rise of almost real-time communication channels (such as RSS-based monitoring of content, support for comments by many content sources and microblogging solutions), and propagation of more informal expression forms (even in public communication).
5. Moving from simple to complex activities. In the early days of the Web, typical users focused mostly on browsing content provided by other people. With progressing "read/write Web" philosophy users became involved in more interactive activities such as commenting or tagging content. However, it is only recently that crowds of users has started to be involved in much more complex activities such as on-line multimedia designing, crowdsourced R & D[50] or collaborative ontology development (in explicit way as in [28] or in implicit way, based on other collaborative actions as in case of [22]).

5.2. Areas of Life Altered by the Web. With development of the Web the part of our life activities that can be (at least partially) performed on-line significantly widened. At the same time, with constantly growing population of Internet users and expanding range of on-line functionalities, it is hard to imagine areas of life that have not been altered by popularization of the World Wide Web.

The Web has for example significantly changed both emotional and physical aspects of relationships, partnership and intimacy. Popularization and always-on mode of instant messaging and different methods of cheap on-line voice and video communication, changed the way people keep in touch with their spouses or partners, both during work hours and free time. At the same time, these communication methods support relationships between people spending a lot of time in distant locations. In parallel, common on-line activities such as Web-based sharing of artifacts (photos, music, links to interesting articles etc.), participation in on-line games or 3D worlds, exchange of digital gifts or collaborative creation, are becoming an important part of shared experiences of many contemporary couples. Finally, Web-based dating services and social networking sites support also formation of relationship, enabling search for partners both for long-lasting relationships based on romantic love and partnership, as well as for short-term, often sex-oriented relationships.

Similar changes are happening even more intensely in the area of friendship and social life. WWW enables easier participation in multiple social groups, varying from communities of practice and domain experts discussion forums, through different forms of on-line activist, charity or political communities, to various on-line multi-player games or virtual words fans. Some of such on-line groups bring together people with very specific interests, that are shared by few of their off-line colleagues, thus inciting the strong sense of belonging. In other groups the connection of members is strengthened by off-line activities they perform together. The Web has also a very significant impact on people reputation and status, because it works like archive of large part of our social activities. This impact is limited not only to what a person did or said on-line, but also on-line gossip or word-of-mouth about him/her. Moreover, the impact of on-line reputation is not limited to on-line activities. More and more companies skim through social media services while recruiting new employees. It is to be noted that the importance of on-line status and reputation is one key drivers of a number of collaborative efforts such as knowledge-exchange forums, open-source communities or Wikipedia (with expertise-based status), and social networking sites (with number of connections being one of elements of status).

While the education systems tend to adapt slowly to progressing "internetization" of our lifes, education, self-development and socialization has been significantly alerted by on-line services. They totally changed the way one can acquire information, thus engendering need for capabilities related to filtering, understanding and merging facts from multiple sources. The philosophies of distance and life-long learning became more feasible

[50] Examples include system that support design (e.g. in footwear or t-shirts companies such as Threadless and RYZ), and systems that support management of different product and service ideas (with examples coming from Dell, Starbucks and Salesforce; see: `http://www.readwriteweb.com/archives/ideascale_launch.php`).

because of development of e-learning (both involving teachers and using solely on-line resources). A plethora of e-learning solutions gives students more interactivity, better adaptation to their learning style and different ways of team learning. Moreover, as the Web is a very information-intensive space, even its everyday usage may be considered a type of self-development experiences.

The development of the Web have huge impact not only on development and socialization of oneself, but also on parenthood, i. e. on socialization of children. One one hand it is an extensive source of information and a communication channel joining with other people with parenthood experience (which is especially important for people with specific problems with their children such as rare diseases). On the other hand, the Web with its advantages and dangers is one of topics that need to be handled by parents in socialization process. The importance of wise parental education in this area is constantly underlined by a number of social groups both supporting IT capabilities development in children, and fighting different types of on-line abuse.

Majority of already described changes are reflected in the way people work in modern organizations. Information sources, methods of contacting other employees, customers and other business entities, approaches to sharing knowledge as well as the percentage of time one works on-line (including partial or even full-time tele-work) has changed dramatically in recent years. The arrival of Web 2.0 and Enterprise 2.0 functionalities to more and more companies[51] suggests that this revolution is not over yet. On the other hand, the symmetric change in how people interact with companies providing goods and services can be observed. It concerns the way we select products (changing mostly due to more accessible and searchable information, better price comparability and wider access to user experience stories), the way transactions are performed (on-line orders, Web-based providers contact, Web-based access to digital goods and services) and the way support and maintenance is delivered (on-line manuals, Web-based support, access to other users community, downloadable updates of software and firmware).

The Web has also a direct impact on ways of spending free time. It promotes a number of on-line individual and social hobbies (such as watching videos or playing on-line games). It supports off-line hobbies by giving wider access to information and to communities of people interested it. One of free time areas that are influenced in these both ways is related to accessing cultural heritage. The Web supports both on-line culture access (by providing virtual museums, live concerts and by including famous places in virtual worlds), and gives information about possible off-line activities and specific cultural items. The Web changes even the very conservative areas related to religiousness and spiritual life. However in this case it typically influences solely information-seeking activities. Finally, it is also an extensive source of information regarding physical activities, health and fitness.

5.3. How the Web Changes Social Landscape. As mentioned before, the Web has impact on almost all areas of life. While the Web introduces some brand new trends to the social life, in majority of cases it just strengthens the tendencies observed previously.

Following the changes that happened in 19th and 20th century, the Web significantly enlarges the space of choice in all areas of life. It gives access to an enormous amount of information concerning products and services, religions, hobbies, attitudes, cultural goods and people. It implements more and more complex search and recommendation facilities for access to these information. Finally, it enables communication with people all around the world and participation in (not necessarily geographically-bounded) niche communities focused on specific topics or activities. As a result, instead of participating in one society with single, imposed culture, semantics and values, one can select to interact with a number of specific social groups with different—possibly conflicting—perceptions of the world.

Such a change also supports further increase of importance of achieved status as compared to ascribed status. Many components of ascribed status (sex, race, health condition) are invisible or mostly invisible on-line. On the other hand, in large on-line communities (such as open-source community [29], forums or on-line auctions) user's social status and reputation (often measured automatically, based on past interactions) are one of basic measures of trust in given user.

These tendencies have impact on growing complexity of identity of Web users. The Web enables users to have multiple roles, participate in a growing number of groups (the notion of "neighborhood" is redefined by the Web), define herself through participation in different social networks. Moreover, the Web also gives possibility of separation of identity and person. Single user may have different (not connected) and not necessarily fully truthful nor connected to real personal data identities in multiple Web sites, making user profiles a part of "impression management, self-presentation" [9]. These tendencies to purposefully construct selves, altogether

[51] See for example: `http://www.readwriteweb.com/archives/study_fast_growing_us_companie.php`.

to mostly verbal and visual format of information transmission on the Web, make Web identities exceptionally well interpretable on a ground of symbolic interactionism theory.

Way of defining and describing self on the Web was much simpler and more limited only few years ago. People shared information about themselves mostly by constructing homepages and by using signatures in e-mail, discussion groups and Usenet (with GeekCode[52] being example of concise description of some aspects of person with a restricted, concise language codes). Since then, we have observed a number of new methods of expressing self. Many home pages have evolved into Weblogs, typically offering more frequent and time-determined update about given person. Additionally, the trend to create multiple profiles in different kinds of social services (including social networks, people catalogues and search engines, gaming Web sites, dating services and corporate Web sites) is growing in importance. Many of contemporary profiles provide information encoded at least with some level of semantics (varying form lightweight and widely used microformats to fully-fledged but still rare ontology-based representations), making machine processing of the profiles a much easier task. Moreover, new profiles are now easier to create and to be connected to existing profiles thanks to technologies such as OpenID and trends related to social network and social data portability. Other tendency is related to popularization of egocentric social networks, that help define self through social position or family and friendship relations (including also relations to fake identities such as pop culture icons). New quantitative, activities-based components of on-line identities evolved in the context of Web forums (where number of posts or average score of posts became part of social status definition), specific communities (e.g. activity measures in open-source community) and electronic commerce (where quantity and percent of positive comments on previous transactions is used as a measure of trust). Finally, growing popularity of microblogging services such as Twitter introduced new, much more dynamic patterns of "continuous self-expression".

Both collective and individual identity in Internet era are much more matter of choice that some time ago. However, the Web also influences people identity indirectly by reinforcing two previously observed tendencies related to socialization: socialization by media and socialization by peer group. The Internet partially takes over the role of both traditional media and extended peer group. Its role as a medium is concerned with dynamic flow (including on-line word-of-mouth [30, 21] and viral marketing [4]) of ideas, symbols, themes and fads (collaboratively referred to as memes) within social communities and crowds. Thus, the Web acts as a catalyst of memetic processes and bottom-up popular culture development, influencing attitudes of Internet users. As observed in [30], "compared to traditional WOM, online WOM is more influential due to its speed, convenience, one-to-many reach, and its absence of face-to-face human pressure". However, this means also more quickly changing ideosphere, reflected in more dynamic and unstable user identities.

This phenomenon concerns also politics, leading to what Yochai Benkler calls networked public sphere [3] and what other have covered by buzzword of "Citizen 2.0"[53]. On one hand such on-line public sphere means that voters use much more different information sources to formulate their opinions and judge individual candidates. On the other hands, people with similar political sympathies tend to group together and to actively take part in electoral campaign. The power of on-line political communities were demonstrated by recent US elections as the victory of Barack Obama was attributed (among others reasons) to his greater on-line activity and presence in social media[54]. While on-line political activities enable better information access and public discussion, they also increase risk of manipulation thanks to personalization of message (based on both: greater possibilities of targeting, and increased number of possible communication channels and formats) or even by hard-to-cease circulation of false, defamatory statements (few examples are given in [19]). Moreover, when people are too involved into communities sharing exactly the same opinions, the real pluralism of thought is replaced by so-called "plural monocultures", inhibiting public discussion.

Contemporary World Wide Web gives users also countless possibilities of expressing themselves in more creative ways, by democratizing social institutions related to culture creation, as well as to values and attitudes promotion. With low cost and high accessibility of media production (including both textual content and simple multimedia), the Web became an oasis of amateurism with amateur actors, performers, writers, directors and editors. While it means more freedom of creation, it also makes it harder to sieve through tons of unverified content. In general, it also means that free, amateur and dynamic content replaces at least some part of paid, professional, static and verified content.

[52] `http://www.geekcode.com/geek.html`.
[53] See `http://www.slideshare.net/jessesaves/citizen-20/` for an overview.
[54] See for example: `http://www.readwriteweb.com/archives/social_media_obama_mccain_comparison.php`.

Over centuries we observed the growth of accessibility of information on other people life. It is partially because of cultural changes that make many aspects of life come out of taboo, but it is also technological progress (especially in communication technologies) that made it simpler to acquire information on others directly from them (compare letters brought by horses, air mail and phone). With popularization of social on-line services this tendency was brought to a new level. Instant messaging, e-mail, social networks (with update tracking), blogs and microblogs allow us to track multiple aspects of life of other people in real time without even a need for direct contact. As demonstrated by research social networking tools mostly help to maintain existing off-line relationships [12]. However, they enable people to keep current about much higher number of friends. This trend combined with aforementioned "continuous self-expression" leads to a phenomenon called continuous partial attention[55], with people attention continuously split between a number of activities, resulting from willingness "to be busy, to be connected, is to be alive, to be recognized, and to matter".

The trends described above are to a large extent continuation of previous social evolutions, with significant quantitative changes related to number of participants, frequency of contact, number of choices we have, or number of impression management channels. These changes are basically quantitative, but with large increase/decrease in numbers, they become in fact qualitative. For example, on-line word-of-mouth is based on graph-like structures that mimic on-line gossip. However, as on-line networks may have much more connections and content or ideas spread much quicker, the message amplification happening on-line is qualitatively different to off-line social phenomena.

Apart from extension of existing tendencies, a few new social phenomena inherent to the World Wide Web, may be observed. In the history some components of social interaction (such as social structures, networks and expectations) used to happen "behind the scenes"—they had direct impact on life of people but remained hard to observe and understand. With recent developments in the Web, some algorithms became a new element of "behind the scenes" of social interaction. It is through algorithms (often not known or known partially) that search rankings are determined and trust is measured. There are also different kinds of algorithms that suggest what products we could buy (e.g. in contextual or behavioral advertisement), what content may interest us (e.g. in collaborative filtering, personalized search results ranking or in adaptive e-learning systems) and which people would be best partners (e.g. in dating services) or friends for us (e.g. in social networking sites). Moreover, some algorithms generate new content based on some kind of statistical or logical reasoning. For examples, automated summarization and aggregation of user opinions is used as a generalized perception of specific products or businesses[56], analysis of news and value of neighborhood houses may be used in valuation of real estate[57], and natural language processing and information retrieval technologies are used by people search sites to construct people profiles from multiple dispersed facts[58]. In all of these cases, the algorithms have direct impact on perception of products, businesses, real estates and people by on-line users. Finally, many types of algorithms performing business activities (e.g. trading algorithms used in stock exchanges or "sniper" software used in on-line auctions) shape contemporary economic environment.

5.4. Paradoxes of contemporary WWW. Some changes happening on the Web have a rather paradoxical character. On one hand we observe the socialization of previously unsocial phenomena—we observe for example the democratization of creation and growing social control over media. On the other hand many activities that are clearly social off-line may be performed in partially "dessocialized" way on-line. For example many of on-line communication channels enable anonymous discussions, and some dating services implicitly support short-term, no-involvement acquaintances (often with people providing at least some fake personal information).

On one hand the Web is the space of almost unlimited choices that enables in a much more flexible way to "be oneself" both in terms of individual self and collective identities of niche communities (the Web enables preservation of folklore or specific languages, supports contact with once culture even in foreign countries, supports development of niche, "long-tail" products, media, services and communities). On the other hand it promotes uniformization at the unprecedented level, if you are not determined enough to build up your identity. WWW is strongly dominated by only a few languages and supports quick propagation of cultural patterns and

[55] See: `http://continuouspartialattention.jot.com/WikiHome`.
[56] For example Pluribo (`http://www.pluribo.com/`) automatically summarizes Amazon product reviews.
[57] For example in case of Zillow, `http://www.zillow.com/`.
[58] Examples include Pipl (`http://pipl.com/`), Spock (`http://spock.com`) and PeekYou (`http://www.peekyou.com/`).

ideas, often driven and supported by commercial activities. That is why Internet popularization programs such as *One laptop per child*[59] are accused of cultural colonialism.

On one hand, many contemporary on-line services support user creativity through discussion, modification, combination or reorganization of existing content. On the other hand, many of such activities are limited to rather mechanical and uncreative (not to say unthoughtful) copying and pasting of information, crossing the thin line between creative combination and plagiarism or generation of noisy, not understandable content.

Finally, the Web is the space of contradictory developments regarding professionalization and amateurism. On one hand, we observe a flooding of amateur content in all areas of the Web (including such capital-intensive areas as film-making[60] and such traditionally restricted areas as legislation[61]). On the other hand, the Web is the major source of income of a growing community of professionals, and we observe a continuous professionalization of technologies handling information collection, processing and search.

6. Inter-component Dynamics. The previous section analyzed separately the changes happening in four components of contemporary Web: infrastructures, content and data, functionalities, and usage. While such abstraction allows easier understanding of some processes, the forces that span multiple components needs to be profoundly studied.

6.1. Demand-driven vs Supply-driven Developments. One of most interesting questions related to inter-component dynamics are related to causality and driving forces behind the observed large-scale changes. During our analysis we identified two opposite—yet complementary—sources of motivation for formation of complex on-line systems and usage patters, leading to demand-driven and supply-driven developments.

The former consists in a series of requirements-driven relationships. Real needs of user and business are the ultimate condition of success of new proposed approaches. Thus, they have direct impact on proposed functionalities, which define requirements of both information representation methods and basic infrastructures. This is the way that majority of on-line services were created. Search engines and Web directories (with underlying infrastructure) were created to enable easier information access, Usenet and e-mail for communication purposed, peer-to-peer solutions aim at enabling easy file sharing (disregarding copyright regulations), content and presentation separation (e.g. HTML+CSS or XML+XSLT) simplifies Web page and Web applications development, RSS aims at keeping visitors current about Web site udpates, and OpenID is supposed to simplify logging into multiple services.

The latter relation is supply-driven and acts in opposite direction. Infrastructural developments lower the barriers for new forms of content representation and types of services. At the same time, better information representation supports development of more sophisticated functionalities which in turn may create new needs and habits of users, as well as new business models. As a result, the endless possibilities of combining existing and new resources, functionalities and user actions enable new, creative, complex on-line services. Thus, a number of services is just a by-product of some need-driven developments. For example infrastructure developed for Web indexing or other large-scale Web applications, promoted gigabyte size mailboxes and a development of cloud-based applications (many of which never could afford enough IT infrastructure in no cloud solutions existed). Search infrastructures enabled also to observe what information is accessed by people, allowing for example to detect flu outbreaks[62]; at the same time, they made possible large-scale empirical Web studies without own crawlers. Usenet and e-mail were successfully used to transfer large files (through peer2mail services), many peer-to-peer solutions are now used in fully legal content distribution or in VoIP communication (well-known example of Skype), XML and RSS technologies enable a myriad of services combining content from multiple locations, and OpenID makes it much easier to collect information about single person from multiple social Web locations.

These two directions are strongly complementary and support one another. Users' demand incites creation of new infrastructures, information representation methods and service interaction models (demand-driven direction). However, once they are created, they pose an opportunity for development of new services (supply-driven direction). Moreover, new services often modify users and businesses perception and engender new needs, that start another wave of innovation.

[59] `http://laptop.org/`.

[60] First feature film fully created by fans using the Web (via Massify, `http://www.massify.com/`) is planned to premier in January 2009

[61] See for examples: `http://blog.wired.com/27bstroke6/2008/03/stanford-law-pr.html`

[62] See: `http://www.google.org/flutrends/`.

6.2. Impact of Infrastructure Development on Functionalities. As we described in Section 2.2, many specific classes of features are becoming today a domain-specific infrastructure, provided by large players, profiting from economies of scale. This progress typically lowers entry barriers and operational costs for new businesses. Thus, it has positive impact on innovation and on enriches set of functionalities accessible to the users. At the same time, it poses two groups of challenges to existing businesses. On one hand, the smaller companies operating in the areas that get "infrastructuralized" typically are unable to compete with large-scale players and need to provide different kind of value-added. As a result, the whole areas becomes cannibalized by infrastructure operators (see Table 6.1). On the other hand, lower entry barriers and operation costs, as well as changing business models of companies leaving cannibalized areas lead to more aggressive competition and dynamically changing competitive environment, thus limiting expected ROI and increasing strategic risk.

TABLE 6.1
Areas of research and business that may be cannibalized by new infrastructures

Services	cannibalized domains
Yahoo Pipes!	commercial mashup creation tools
OpenCalais	natural language processing software, research in information extraction from text
VoIP solutions	traditional telephony
distributed storage solutions and cloud computing	ISPs
folksonomy-based content organization	Web page directories
automated on-line translation services	professional translation services

6.3. Business Models. Business models are another element of inter-component dynamics. As no sustainable services can be provided on a long term basis without a business model (defining economic feasibility of specific enterprise), they shape the development of the Web at all mentioned levels and between them. Aforementioned infrastructurization of some part of traditional value-added of on-line companies is one of challenges of today business models. However it is not the single nor the most important one.

First area that every business model needs to address is related to revenue sources. Traditional solutions is this area include sales of goods, sales of services, acquiring commission from other businesses and sales of advertisement space. Sales of goods is currently the major source of income of on-line economy in general. However only a limited number of businesses (such as on-line stores, auction platforms or virtual malls) focus on activities related to e-commerce, and another small group of on-line services sell some items (mostly hobby-related) apart from their main operations. Today's e-commerce is shaped mostly by growing accessibility of machine-processable information about customers, competitors and suppliers, better analytical tools (including data mining, business intelligence systems as well as rule-based mechanisms for automation of transactions or other business processes), and outsourcing of non-core activities (with many shops sending goods directly from their suppliers inventories through drop-shipping, and some shops outsourcing all logistics and transaction-related activities as it is in case of Zlio.com shops). At the same time, majority of goods sold on-line become commodities accessible from multiple providers. Together with better information access this trend strengthens price competition. To circumvent this dangerous, margin-cutting tendency many businesses try to provide value-added related to after-sales services (e.g. updates, insurance, warranty, support for switching to new models), combined sales of goods and services (e.g. in telecommunication area) or personalization of products (varying from simple customization of physical product as in case of Fiat 500, through products developed in co-operation with user such as t-shirts or puzzles constructed from user photos, to products that are physically identical, but differ by accompanied services or digital goods).

The area that is supposed to prosper most in years to come is related to sales of on-line services, both computerized and performed manually. While the traditional, subscription-based information services decline and will probably be limited to a series of niche markets (e.g. access to specific databases), we observe a dynamic rise in sales of infrastructural services (e.g. storage, computing, API-based search), different types of services implementing pluggable complex logic (e.g. automatic or semi-automatic translation, accounting, massive sending of paper mail or faxes), services supporting different types of analytical activities (e.g. competitive analysis, market monitoring, search engines optimization) and access to on-line software (sold in *Software as a Service* philosophy). At the same time, a major shift in pricing models can be observed in this area, from traditional

one-time or subscription fees, to fees based on actual usage (e.g. used computing power or storage, number of invocations, set of used software features).

As stated before, we observe a decline of paid information-based services. At the same time, information-intensive Web sites remain among the most popular destinations on the Web. In contemporary Web their revenue sources are mostly based on commissions and sales of advertisement. Commission-based revenues are typical for services that provide transaction-oriented information, such as comparison shopping sites or flight-reservation cybermediaries. All over information-centric sites tend to include different types of advertisement. Changes that happen in on-line ads industry concern mostly support of new type of media (ads embedded in videos, Flash animations or on-line games), popularization of contextual advertisement, better personalization of served ads (with behavioral modeling, and wider access to information about visitors), and different pricing models (with payment for ads becoming more commission-like and dependent on user attaining specific Web site goals such as transaction or registration).

Second key area of business models is concerned with operating costs. In recent years we witness two cost-cutting tendencies related to outsourcing and crowdsourcing. Outsourcing is a business trend for many years, however, recent changes in pricing models (related to pay per usage or pay as you go approaches) and much easier integration with third-party functionalities (which means lower transaction costs and switching costs) made outsourcing more profitable and manageable, in the same time significantly limiting the risk of lock-in. On the contemporary WWW, companies may outsource almost every non value-adding activity, starting with storage, computing and other instrastructural services, and including many technology and business operation tasks. At the same time many activities related to content creation, assessment and organization may be outsourced to the community of users (or crowdsourced) in the spirit of Web 2.0 services. As the users often are not paid at all, paid low wages (being rather a perk than a salary), or remunerated with low-cost, high-value internally produced goods or services (e.g. better or free account, augmented storage quota, higher content modification rights), crowdsourcing may significantly lower costs of multiple activities required by businesses. However, while crowdsourcing key business tasks (e.g. some part of R&D), the companies need to resort to specific quality assurance techniques.

Finally, the third area that has significant impact on all kind of on-line business is related to acquisition and maintenance of user base. In the traditional approach each on-line service aimed at acquiring individually as many users as possible (before competitors can surpass them) and maintaining this user base thanks to network effects and user lock-in. With recent changes related to federated identity (including OpenID) that simplify registration in multiple services, continuous development and professionalization of viral marketing campaigns, accessibility of more social-networked channels (making propagation of ideas and links even easier) and better technologies handling load peaks (e.g. cloud computing) this approach becomes even more feasible. However, experiences from early years of 21st century suggest that huge user base does not guarantee success, underling importance of revenues, costs and clear value-added. Moreover, with progressing tendencies towards data and social network portability the strength of lock-in of both users and business customers is continuously decreasing. Additionally, past experiences indicate that switching costs and lock-in effects should be counted among top criteria for selection of IT solutions. Finally, with such solutions as Facebook Platform, it is also much easier to access huge user bases of existing services. All this tendencies support more organic and value-added-centric growth of audience of on-line services.

6.4. Privacy in the Big Brother's Era. Another area that contains all components of contemporary WWW is related to user privacy. Almost every activity that is performed by people on-line leaves a number of electronic traces. Each server that is involved in complex functionalities (including proxy server and enterprise proxy servers), Web analytics software, search engines and many other services collect data regarding user behaviors. In some cases these data are directly connected to user profiles, in other cases they are anonymous but span multiple sessions and contain a lot of information about specific user (sometimes this information e.g. queries posted to a search engine is satisfactory to identify specific users). Moreover, in case user uses the same profiles (e.g. OpenID) in multiple locations, it is easy to connect behavior data from multiple sites. The integration is also simplified by concentration of many services in hands of a few big players (such as Yahoo and Google) that adopt integrated approach to tracking users. As a result, for example Google may merge browsing sessions of its search engines, all Web sites using Google Analytics, e-mail browsing by GMail, social activities in Blogger and in a plethora of other services owned by Google. Moreover, the rapid progress in Web usage mining and its applications gives the data owner growing insight into how to understand and take advantage of user behavior.

Collection of Web usage data is just one face of personal information on the Web. A lot of activities leave permanent and public results such (micro)blog posts, Usenet or discussion groups messages, comments in multiple forums, or created tags. Many informations are also shared in social networking services (they include not only information on given person but also on people (s)he is connected to), and other location such as user profiles (including home pages, institution pages, university students lists, and information legally required to be public). With growing machine-processability of Web content these information are much more easy to integrate, giving more complete view of specific user's hobbies, opinions, political and organizational involvement and colleagues.

6.5. Towards Ecosystem-Based Computing Paradigms. Many of tendencies described in previous sections of this paper involve complex interaction between multiple objects—complex flows of information between numerous services and people, composite software using multiple independent and dispersed logic components as well as numerous and heterogeneous information sources, complex interaction between multiple businesses resulting from growing outsourcing, and finally interaction between software components and people who may be involved in information flows and provide feedback on algorithms results.

All this tendencies converge, to form complex ecosystems involving software components (algorithms), people (individual acting on their own behalf, individual acting on behalf of organizations, intelligent crowds), different types of content and data, and different types of organizations (represented by business processes, procedures and rules; in case of governments it includes also legislation). This combination can be considered a new Web-based computing paradigm, concerning solving complex problems at the level of social processes.

There are a few characteristics specific for this computing paradigm. First of all, in this paradigm computing is a mixture of machine and human computing. The actual data and control flow is performed by a number of algorithmic "black boxes". Majority of them are automated, but some may be contain manually performed logic, combine manual and automated activities. It is to be noted, that these black boxes, that are composed to obtain complex workflows, may also consist of multiple embedded logic components.

Secondly, in this paradigm computing is an area of constant changes. They concern both changing internal (hidden) logic of components and changing composition of components. For example, while the basic functionality of search engines does not change and the syntax changes rarely (with backwards compatibility), used search algorithms continuously evolve. At the same time, mashups are continuously change, infomediaries and meta-search engines include more information sources, and user-created pipes can be modified within moments (when needed). Moreover, the solutions that change composition of on-line logic according to user or business process needs and past performance of specific services are around the corner. It is to be also noted that, as results of multiple logic components and workflows are stored and publicly available, many complex flow are implicit and not designed by anyone. For example whenever content resulting from some text mining or data extraction activity is stored, it is next indexed by general purposed search engines and can be included in some search-based scientific or market research workflows. All these characteristics result in computing which is distributed not only at the level of computing power (which is assured for example by cloud-based solutions), but also at the level of logic (multiple competing workflows performing similar but not identical activities can be performed in parallel, combined, compared, used to create new workflows). On the other hand, it means that results of such complex flows are not deterministic.

Finally, new paradigm of computing that we observe is not limited to flow of data and control between multiple logic components. Majority of both automated and manual tasks performed on-line have their business context. For example, many activities create legal obligations and cause money flows. On the other hand, business rules—that may depend on internal company conditions—are a an important component of control flow. For example, product search activities may end up by a transaction provided that product price is exceptionally low and company has enough of stock space at the moment of planned delivery.

7. Conclusion. In this paper we presented a bird's-eye view of changes that has happened recently at the WWW infrastructure, resources, functionalities and usage areas, varying from very technical developments to social changes that follow. We started by analyzing separately each of these components of contemporary World Wide Web, and then moved on to dependencies and relations between them. At the final part of this article we shortly presented how convergence of described changes leads to new computing paradigm, combining large variable of dynamically changing logic components with human participation and business perspective.

REFERENCES

[1] A. All and T. Byrne, *Still a big gap between reality, wishes for web 2.0.* http://www.itbusinessedge.com/item/?ci=44015, November 2008.

[2] V. Anupam, J. Freire, B. Kumar, and D. Lieuwen, *Automating web navigation with the webvcr*, in 9th International Conference on World Wide Web, 2000, pp. 503–517.

[3] Y. Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, Yale University Press, 2006.

[4] P. Blackshaw and M. Nazzaro, *Consumer-generated media (cgm) 101: Word-of-mouth in the age of the web-fortified consumer*, tech. report, Nielsen BuzzMetrics, 2006.

[5] U. Bojars, J. G. Breslin, V. Peristeras, G. Tummarello, and S. Decker, *Interlinking the social web with semantics*, IEEE Intelligent Systems, 23 (2008), pp. 29–40.

[6] J. Borges and M. Levene, *Data mining of user navigation patterns*, in Workshop on Web Usage Analysis and User Profiling, 1999, pp. 31–36.

[7] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T. Chandra, A. Fikes, and R. E. Gruber, *Bigtable: a distributed storage system for structured data*, in OSDI '06: Proceedings of the 7th symposium on Operating systems design and implementation, Berkeley, CA, USA, 2006, USENIX Association, pp. 205–218.

[8] M. Cooper, *Accessibility of emerging rich web technologies: web 2.0 and the semantic web*, in W4A '07: Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A), New York, NY, USA, 2007, ACM, pp. 93–98.

[9] danah m. boyd and N. B. Ellison, *Social network sites: Definition, history, and scholarship*, Journal of Computer-Mediated Communication, 13 (2007), p. article 11.

[10] T. H. Davenport, *Competing on analytics*, Harvard Business Review, (2006).

[11] J. Donath, *Sociable media (prepared for the encyclopedia of human-computer interaction).* http://smg.media.mit.edu/papers/Donath/SociableMedia.encyclopedia.pdf, April 2004.

[12] N. Ellison, C. Steinfield, and C. Lampe, *The benefits of facebook"friends": Exploring the relationship between college students' use of online social networks and social capital*, Journal of Computer-Mediated Communication, 12 (2007), p. article 1.

[13] M. Franklin, A. Halevy, and D. Maier, *From databases to dataspaces: a new abstraction for information management*, vol. 34, New York, NY, USA, 2005, ACM, pp. 27–33.

[14] B. He and K. C.-C. Chang, *A holistic paradigm for large scale schema matching*, SIGMOD Record, 33 (2004), pp. 20–25.

[15] J. Hendler, N. Shadbolt, W. Hall, T. Berners-Lee, and D. Weitzner, *Web science: an interdisciplinary approach to understanding the web*, Communications of the ACM, 51 (2008), pp. 60–69.

[16] P. Heymann, G. Koutrika, and H. Garcia-Molina, *Can social bookmarking improve web search?*, in International Conference on Web Search and Web Data Mining, 2008, pp. 195–206.

[17] J. Howe, *Wired magazine: The rise of crowdsourcing.* http://www.wired.com/wired/archive/14.06/crowds.html, June 2006.

[18] M. Hu and B. Liu, *Mining and summarizing customer reviews*, in KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, New York, NY, USA, 2004, ACM, pp. 168–177.

[19] A. Keen, *The Cult of the Amateur: How Today's Internet is Killing Our Culture*, Doubleday Business, June 2007.

[20] R. Kosala and H. Blockeel, *Web mining research: a survey*, ACM SIGKDD Explorations Newsletter, 2 (2000), pp. 1–15.

[21] J. Leskovec, L. A. Adamic, and B. A. Huberman, *The dynamics of viral marketing*, ACM Trans. Web, 1 (2007), p. 5.

[22] M. Z. Maala, A. Delteil, and A. Azough, *A conversion process from flickr tags to rdf descriptions*, in SAW 2007: Social Aspects of the Web, D. Flejter and M. Kowalkiewicz, eds., vol. 245 of CEUR-WS, 2007.

[23] S. K. Madria, S. S. Bhowmick, W. K. Ng, and E.-P. Lim, *Research issues in web data mining*, in 1st International Conference on Data Warehousing and Knowledge Discovery, 1999, pp. 303–312.

[24] L. Manovich, *The Language of New Media*, The MIT Press, March 2002.

[25] J. Park, T. Fukuhara, I. Ohmukai, and H. Takeda, *Web content summarization using social bookmarking service*, tech. report, 2008.

[26] L. G. Reid and A. Snow-Weaver, *Wcag 2.0: a web accessibility standard for the evolving web*, in W4A '08: Proceedings of the 2008 international cross-disciplinary conference on Web accessibility (W4A), New York, NY, USA, 2008, ACM, pp. 109–115.

[27] F. Rose, *The life cycle of a blog post, from servers to spiders to suits – to you.* http://www.wired.com/special_multimedia/2008/ff_secretlife_1602, January 2007.

[28] K. Siorpaes and M. Hepp, *myontology: The marriage of collective intelligence and ontology engineering*, in Proceedings of the Workshop Bridging the Gap between Semantic Web and Web 2.0 at the ESWC 2007, LNCS, Springer, 2007.

[29] D. Stewart, *Social status in an open-source community*, American Sociological Review, 70 (2005), pp. 823–842.

[30] T. Sun, S. Youn, G. Wu, and M. Kuntaraporn, *Online word-of-mouth (or mouse): An exploration of its antecedents and consequences*, Journal of Computer-Mediated Communication, 11 (2006), p. article 11.