Investigating the Suitability of Web X.Y Features for Software Engineering – Towards an Empirical Survey

Eric Ras, Jörg Rech, Sebastian Weber

Fraunhofer ISE
Fraunhofer-Platz 1
67663 Kaiserslautern
Germany
eric.ras@iese.fraunhofer.de
joerg.rech@iese.fraunhofer.de
sebastian.weber@iese.fraunhofer.de

Abstract: Today, software engineers strongly rely on information while they perform development activities in the different software engineering (SE) phases. The results of a previous survey showed that most information is required during the phases of requirements engineering (RE), design, programming, and project management. Web X.Y features (i.e., concepts and technologies) facilitate collaboration and communication with distributed individuals and help to cope with the immense amount of information by simplifying the organization, integration, and reuse of information scattered across diverse content sources. After presenting the features of the different Web X.Y generations (i.e., Web 2.0, Web 2.5, and Web 3.0), we propose a weighted mapping for the relevancy of these Web X.Y features regarding their support for the four SE phases with the highest need for information. Based on this subjective mapping, a set of research questions and hypotheses is derived that form the basis of an empirical survey. The goal of this survey is to investigate the potential of Web X.Y features for SE.

1 Introduction

The work environment in software engineering (SE) is changing rapidly, as new methods, techniques, and tools keep popping up to cope with the rising demands of customers asking for high-quality software. Nowadays, software engineers face an immense amount of information that is necessary to perform daily engineering tasks. Requirements and technologies are always in flux, legacy software needs to be understood and changed, the architecture and design need to be adapted, and test results or changes need to be communicated within the team. Web X.Y features (i.e., concepts and technologies) provide new possibilities for low-threshold, lightweight mechanisms for supporting the search, access, and usage of information during the different SE phases.

In this paper, we present a mapping between Web X.Y features and SE phases with high information needs, because here, a high potential of Web X.Y concepts and technologies...
can be expected (Section 2). While concepts such as collaboration, technology-based knowledge sharing, and ubiquitous access to information are essential in all development processes, we focus on the traditional SE lifecycle, although global, model-driven, or open-source software development could benefit from Web X.Y concepts and technologies as well. One of the main contributions of this paper is to propose a weighted mapping of the relevancy of Web 2.0, Web 2.5, and Web 3.0 features (see Section 3 for their descriptions) regarding their support for a selection of SE phases based on judgments made by researchers. This mapping is explained exemplarily for the phase of requirements engineering (RE) in Section 4. A second contribution is a set of research questions and hypotheses, which serves as a basis for developing an empirical survey. The aim of this survey will be to investigate the information need, information usage, collaboration behavior, and knowledge exchange in specific development phases as well as the general potential of Web X.Y features in SE. Finally, Section 5 summarizes the paper and describes the next steps.

2 Information Needs in Software Engineering

SE in a software organization have special needs for information in their daily work that are not supported by most current search methodologies and technologies. During most of their working time, software engineers go through an incremental process of acquiring, evaluating, organizing, analyzing, presenting, and securing information [Do09]. Achievements used in Web X.Y systems offer new opportunities for connecting and supporting these software engineers by exploiting network effects as well as the information available in a software organization.

In a first step, we conducted a survey on intelligent assistance in SE to find out where most information is needed and which kind of information resources are used in software organizations [RRD07]. We discovered that software engineers are often confronted with new application areas and technologies requiring them to search for suitable information. We further investigated the respondents’ retrieval rationales and present the results in Figure 1. It shows that the main learning rationales are to solve concrete problems at hand and close knowledge gaps, and personal motivation.

Furthermore, Figure 2 depicts phases in SE where the participants frequently need information or support by colleagues in order to make decisions, solve problems, or enhance their competencies. The data shows that the early development phases such as project management, RE, software design, and programming are the phases with the highest need for additional information.

In addition, we investigated which information sources are used at the workplace. Not surprisingly, most software engineers are using Internet search and personal contacts as their sources of information. In addition, company-specific knowledge bases and help systems in the applications used are utilized to satisfy their information needs. Regarding our three differentiation groups (experience, position, and company size), we could not find any significant differences in the use of information sources.
Next, the different Web X.Y generations and their features are explained in order to simplify the mapping of Web concepts and technologies to those generations.

![Figure 1: Rationales for information gathering](image1)

![Figure 2: Information need in software engineering phases](image2)
3 Web X.Y Generations

The World Wide Web (WWW) has been through many changes since its beginnings and has become the largest information platform worldwide. In the following, we describe the evolutionary steps of the Web, i.e., Web X.Y generations and their features (concepts, such as folksonomies, and technologies, such as blogs) – the main aspects that characterize the particular Web era. While new generations of the Web supersede older ones, concepts and technologies from new stages do not completely replace older ones but co-exist with them (e.g., email, FTP, blogs, etc.). However, in newer stages, concepts or technologies might be replaced by advanced solutions (e.g., feeds in the Web 2.0 stage might be replaced by intelligent assistants in the Web 3.0 stage). While the term “Web 2.0” goes back to its identification by O’Reilly [Or09], Weber and Rech proposed definitions and classifications of the terms Web 2.0, 2.5, and 3.0 based on a systematic literature survey as briefly described throughout the next sub sections (see also [WR09]). During the initial Web step (Web 1.x), Tim Berners-Lee developed the fundamental architecture of the Web. All the inventions and concepts of this phase (e.g., HTTP, linking documents, or referencing Web resources by a unique URI) are still integral parts of the Web.

3.1 WEB 2.0 Features

In Tim O’Reilly’s [Or09] famous essay, he described Web 2.0 as a new stage in the evolution of the Web. In the spirit of Web 2.0, Web-based applications make the most of the intrinsic advantages of the Internet as a platform. They get better as more people use them by capturing network effects; they enable collaborative work; they deliver rich user experiences via desktop-like interfaces; and they combine data from multiple sources into new services.

Social Networking software (e.g., LinkedIn) is a special kind of social software that supports human communication and collaboration [St05]. Social networking services offer all or some of the following exemplary functionalities: network of friends, person browsing, private messaging, commenting, etc. The main benefits are that users can interact, exchange information, and keep updated with their networks.

Annotating refers to activities where users create content or comment on existing content. This is a common communication concept for collaboratively creating content and improving the quality. Rating is a special kind of annotation. Users can assess content in order to communicate whether it was relevant or not to them for the current context.

Web applications featuring recommendations suggest information originating from social communities (e.g., Amazon customers) to their users in order to increase the benefit.

Mashups refer to an ad-hoc composition of content and services coming from different sources (e.g., Websites, RSS feeds, or databases) to create entirely new services that were not originally provided by any integrated source.

In syndication, the content’s essentials are, in effect, lifted from the traditional website.
context and published without styling information but with enhanced metadata (title, date, links, etc.), which makes it possible for end-user tools to mashup the content with material from other sources [Ay06]. A widespread syndication technology is RSS.

**Folksonomy** is the practice of collaboratively creating and managing tags to annotate and categorize content [Vo07]. The basic principle is that end users do subject indexing instead of experts only, and the assigned tags are shown immediately on the Web. This mechanism can help to find content and to collaboratively create a common vocabulary.

**Collaboration** refers to Web applications that enable the users cooperatively create or annotate content. Examples are enterprise wikis where project experience is documented.

The **blog** phenomenon impressively reveals that Web 2.0 has brought a shift in the way people use the Web. Expressing one’s own interests instead of “blind” consuming and staring is the motto. Blogs are simple “Web diaries” that can be created by anyone thanks to simple tools. Track-back links and comments make this concept interactive.

**Life streaming** refers to an online record of a person’s daily activities, either via direct video feed or via aggregating the person’s online content such as blog posts or social network updates [Ws09].

### 3.2 WEB 2.5 Features

Web 2.5, the Pervasive Web, will focus on users who are “always-on”, carrying along their mobile devices connected to the Internet. There is a shift away from “desktops” as unique Internet access towards increased usage of mobile devices – off-site reading (e.g., with RSS feeds and Web desktops) and publishing (e.g., Twitter as a microblogging service or Diigo as a social annotation service) will be integral parts of Web 2.5 technologies, which constitutes the transition between Web 2.0 and Web 3.0.

**Seamless Experience** refers to the fact that multimedia, games, or normal content can be seamlessly viewed or played on multiple devices without remembering positions, replaying levels, or changing formats for the specific device. The devices change to an ambient system that can be accessed anywhere and anytime (e.g., at home, in the car, or at work).

**Device Sensitivity** means that services and content are becoming sensitive to the device they are viewed on. Interfaces and functionalities adapt to the input or output capabilities.

**User Sensitivity** means that services and content are sensitive to the user (profile) that operates or views it, such that they are presented in understandable languages or in different colors if the user is color-blind.

**Location Sensitivity** means that services and content are sensitive to the location they are viewed at, e.g., when micro-blogging tools are disabled in cinemas or ads are presented in the context of stores located in the vicinity.

**Content Sensitivity** means that services are sensitive to the content they present, such as job postings being offered if a CV is designed.
Time-Sensitivity means that services and content are sensitive to the time or season they are viewed at, such as navigational devices being darkened or the hint about a special lunch offer on mobile devices only appears at lunchtime.

Dynamic Mashups refer to a dynamical adaption of the data, functions, users, or presentation components available, such as the most relevant blogs being computed dynamically based on the current user profile and integrated into one personalized news stream.

Rule-based Services are simple intelligent features that automatically support the user with information or services. These may range from simple exclusion rules (e.g., parental guides or filters for explicit content) to software agents with different strategies for acquiring products from online auctioning systems.

End-user programming: Since most stakeholders do not have the time or inclination to learn the tools and skills of a professional programmer, they can profit from end-user friendly tools (e.g., visual programming) that enable them to easily assemble context-specific and short-dated “micro applications (service mashups)”.

Semantic Social Networks are the next logical step beyond the current social networking technology as more (semantic) information is available about individuals, which can be used to support more intelligent assistance and pro-active information delivery (PID).

3.3 WEB 3.0 Features

Web 3.0 is often associated with the Semantic Web [LH07], [Ay06], [He08]. Market analyst Davis expects that semantic technologies will embrace all semantic techniques and standards that can be applied on top of Web 2.0 (e.g., basic reasoning or model-based inferencing) [Da09]. According to Murugesan [Mu07], Web 3.0 will make use of already matured Web 2.0 features but also of technologies evolving in Web 2.5. Nova Spivack (2006) defines key emerging technology trends for Web 3.0, such as ubiquitous connectivity, open technologies (e.g., open APIs, protocols, or data), open identity, and the intelligent Web (e.g., Semantic Web technologies, or machine learning) [Sp09]. In the following, we present the Web 3.0 features that emerged from our literature survey [WR09].

Semantic Data refers to the increasing enrichment of content and profiles with additional metadata, which is machine-readable and usable for better differentiation and reasoning.

Semantic Services will build upon semantic data (e.g., using microformats) and offer intelligent assistance to the users. Services exploiting semantic data will provide more intelligent services such as more exact searches or sensitive presentation of data.

Natural-Language Search refers to the ability of search engines to answer full questions such as “Which US Presidents were assassinated?”

Experience-based Services refer to services that learn from the specific behavior of the user and optimize their services in this regard.

The 3D Web has the potential to change collaboration in distributed software projects. Although face-to-face meetings can never be replaced by any technology-based tech-
nique, sometimes they are not possible because of temporal or financial reasons. The 3D Web can blur the boundaries between reality and virtual reality in terms of the next level of virtual meeting rooms, telework, holograms, etc.

4 Potential Use of Web X.Y Features in Software Engineering Phases

In this section, a weighted mapping for the relevancy of these Web X.Y features regarding their support for the four most relevant SE phases (i.e., project management, RE, software design, and programming) is given. Based on this subjective mapping, a set of research questions and hypotheses is derived that form the basis for an objective empirical survey.

4.1 Relevancy of Web X.Y Features for Software Engineering Phases

Table 1 shows the average level of relevance, with the character “○” representing low, “●” medium, and “■” high support. For example, the feature annotation/rating is highly relevant for the phase of RE, whereas for project management, it is only of low relevance. This weighted mapping is based on the judgment of five researchers who were knowledgeable about Web X.Y technologies and SE in general.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Project Management</th>
<th>Requirements Engineering</th>
<th>Design</th>
<th>Programming</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syndication (e.g., RSS, Mashup)</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Social Networking</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Annotating / Rating</td>
<td>○</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Mashup</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Folksonomy</td>
<td>○</td>
<td>●</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>Recommendation</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Life Streaming</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Collaboration (e.g., Wiki)</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Blog</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Seamless Experience</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Off-Site Commenting</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Device Sensitivity</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>User Sensitivity</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Location Sensitivity</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Content Sensitivity</td>
<td>○</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Time-Sensitivity</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Dynamic Mashups</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Rule-based Services</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>Semantic Data</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>○</td>
</tr>
<tr>
<td>Semantic Services</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>○</td>
</tr>
<tr>
<td>Natural-Language Search</td>
<td>●</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Experience-based Services</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>The 3D Web</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
Due to the limited space in this paper, a detailed description of each entry in the table cannot be provided - this would be part of the survey report. Thus, in the following, we will go only into some examples by focusing on the requirements engineering phase and Web 2.0 features because most evidence of the relevancy rating can be provided here due to past studies with Web 2.0 technologies in the RE field (see, for example, [De07]).

4.2 Example: Web 2.0 Features Supporting Requirements Engineering

The RE process can be seen as a creative process that is executed by a requirements engineer. Figure 3 depicts the RE activities graphically. It can be seen that the requirements documentation activity is a core activity located between discovering, analyzing, and negotiating requirements and the validating/verifying requirements. The activity produces the requirements specification that is agreed upon by the different stakeholders. Different types of information are needed to produce the document: stakeholder needs and constraints, organizational standards, regulations, domain information, market trends, and existing system information. Hence, RE is a highly collaborative activity requiring many different types of information from different sources and stakeholders with different preferences, needs, and knowledge. While different approaches to RE exist, which sometimes require formal specifications, many of them can be integrated into Web X.Y tools. For example, Decker et al. used textual and graphical UML diagrams, such as use cases, within wikis [De06].

Requirements elicitation: This activity may require interviewing stakeholders or observing real-world processes. Syndication can first improve the access to information from different sources (e.g., interview questionnaires for developing new interview questionnaires, accessing protocols/recordings/observations from conducted interviews, etc.) by hiding style information, which makes mashing of information possible. Second, syndication can provide more homogenous access to existing systems, work instructions, internal process instructions, and, for example, past change requests. Annotation is a supporting feature for analyzing interview protocols collaboratively. Technologies such as a Wiki can support the collaboration between the different stakeholders, e.g., by supporting brainstorming activities [Ue08]. As stated previously, a lot of information is accessed and analyzed to derive requirements. Folksonomies help to annotate and structure this information and hence enable tools to structure and search in this pool of information.

Requirements analysis: Analyzing means that gaps in the preliminary requirements documentation or conflicting requirements (e.g., contradictions to regulations or standards) are identified or that incompatibilities with constraints such as budget or application environment are discovered. Syndication can enhance access to these regulations or standards and to domain- and customer-specific constraints (e.g., costs, budget, safety constraints). Annotations help to add and semantically connect information, e.g., about potential conflicts with the requirements that have to be discussed with the team. These relationships (i.e., connections) can then be used to perform consistency checks with regard to completeness and correctness (e.g., checking whether each use case description is related to an actor description) [Ue08]. Collaborative commenting supports simultaneous reviewing of requirements between the different stakeholders.
**Requirements negotiation:** If conflicts exist in the requirements, they might be contradicting or incompatible with existing constraints. Resolving these conflicts is the goal of the requirements negotiation process, which can make use of technologies supporting collaboration (e.g., wikis). Prioritizing is also a part of requirements negotiation. It can profit from rating, where the different stakeholders rate the requirements, e.g., according to their complexity, effort needed to implement a requirement, or related risks.

**Requirements documentation:** Requirements need to be documented, which implies that documentation standards and specification rules have to be followed. Again, syndication helps the software engineer to access the multifold types of document guidelines and specification rules. Systems supporting collaboration help to produce a first draft of the requirements specification, where stakeholders with different background and expertise can contribute their knowledge. Hence, requirements can be described more adequately in terms of the stakeholders’ needs and conceptions (e.g., regarding understandability).

![Figure 3: Activities in RE (based on [HS97])](image)

**Requirements validation and verification:** Requirements validation checks if the requirements document expresses what the stakeholder actually wants, i.e., checking whether the document is correct, i.e., complete, without any inconsistencies, agreed upon by all stakeholders, etc. Requirements verification means checking that the specification document fulfils the quality criteria for requirements (i.e., they are formulated clearly, unambiguously, and meet the subsequent development phases’ needs). **Collaboration** and **annotation/rating** enhance the reviewing of requirements (see also [Ue08]).

**Requirements management:** In many cases, requirements change over time. Hence, managing requirements means observing them throughout the development process, i.e., recording and analyzing their changes, tracking their usage in later development phases, and tracking their status. **Syndication** can provide immediate access to changing requirements and can automatically infer other SE artifacts that are impacted by changes too. Software engineers can be informed about the status of requirements and new versions or releases of requirements documents by using, e.g., RSS feeds. **Blogs** are a technology facilitating the documentation of experiences with the implementation of specific requirements and sharing with other developers. It helps to document newly available features or broadcast improvement suggestions or experience with a specific customer.
4.3 Deriving a First Set of Research Questions and Hypotheses for the Survey

As the rating in Table 1 is based on few expert judgments, its reliability and validity are rather low and should therefore not be used as a reference. Thus, our main focus is on deriving a set of research questions and hypotheses to serve as a basis for developing an objective empirical survey. The aim of this survey will be to investigate the information need, information usage, collaboration behavior, and knowledge exchange in specific development phases as well as the general potential of Web X.Y features in SE.

The target group of the survey are primarily SE professionals. This makes it more challenging, because we cannot expect them to also be familiar with the terminology for Web X.Y features as introduced previously in Section 3. Therefore, we need to find a more understandable wording for the questionnaire. Another issue is not to focus on a specific lifecycle model, since this would reduce the sample size dramatically. It is more effective to ask people which lifecycle their organization follows – this would also allow us to investigate differences between various development contexts.

Our previous survey on intelligent assistance revealed that the rationale behind the need for information are users’ personal motivation, solving problems at hand, and closing existing knowledge gaps. In order to investigate the rationale behind the use of Web X.Y features in SE, other dimensions must be added (see also [HMV08] and [MWY06]). Recent studies about experience with Web X.Y technologies in SE could help to improve the questionnaire (RE [Da08], [De07], architecture documentation [BM06], programming [WCM07], project management [Xu07], and processes management [Kr09]).

Besides determining the information need, information usage, collaboration behavior, and knowledge exchange in specific development phases, we are interested in the following general research questions:

- What are the rationales for using specific Web X.Y features in SE?
- How relevant are Web-based collaboration, social networking, and content production as well as consumption for daily SE work in general?
- How relevant are annotations and recommendation of information in SE?
- How relevant is a technical means for mashing up information from different sources (e.g., building an SE dashboard)?
- How relevant are services that are able to deliver information according to the current context (i.e., device-, user-, location-, content-, time-sensitivity)?
- How relevant are intelligent and pro-active services based on Web X.Y features (e.g., inference based on available semantic data, machine learning systems, natural language processing, etc.)?

Based on the mapping between Web X.Y features and a selection of SE activities from Section 4.1, we formulate the following hypotheses, which will be verified or falsified after conducting the survey:

- The average relevancy for blogs and syndication is significantly higher than neutral, independent of the SE phase (neutral refers to the middle value of a (Likert) scale with
a normal distribution).

- The average relevancy of collaboration support is significant higher for RE and project management than for other SE phases.

- The average relevancy of context-aware services (independent of a particular SE phase) is lower than the average relevancy of non-context-aware services.

- The average relevancy of semantic data and semantic services is higher for RE and design than for other SE phases.

- The average relevancy for adaptive and learning systems is higher than neutral.

5 Conclusions and Outlook

Web X.Y features can facilitate communication with distributed individuals and can help to cope with the immense amount of information by simplifying the organization, integration, and reuse of information scattered across diverse content sources. In this paper, a first mapping regarding the relevancy of Web X.Y features for a selection of SE phases was presented. This mapping will serve as a baseline for the development of a survey in the field. A preliminary set of research questions and hypotheses is available to start discussion on which aspects are most interesting to be investigated by such a survey.

The results of this survey will shed light on which Web X.Y features are preferred and should be part of intelligent information platforms for SE in the future. In addition, it will help us to extract, structure, and annotate SE information in such a way that Web X.Y features can use them to provide new information services to engineers.

References


