A Services Theory Approach to Online Service Applications

Claudio Pinhanez
IBM Research, T.J. Watson
pinhanez@us.ibm.com

Abstract

From online multi-player games to web search, the web is now populated by a multitude of online applications. The main question we try to answer in this paper is which of those applications can be characterized as traditional services and therefore benefit from the concepts, techniques, and methods developed in traditional services theory. Our answer is to define online service applications as applications where (1) the user does not control most means of production; (2) the user is a significant part of the input to the production process. We then discuss how some traditional service techniques can be applied to online service applications by considering aspects of the design and evaluation of the human-computer interface (HCI) of online services.

1. Introduction

If we look back into how we were all traditionally taught computer programming, software development, and interface design, we find that four assumptions permeate the understanding of how software is supposed to work: (1) a single user, (2) who controls the machine and data being used, (3) inputting data (4) to be automatically processed by the machine in order to produce some desired information output. In other words, we were taught to see software as a tool owned by the user to be employed when needed.

However, in the last 15 years, with the advent of the WWW and pervasive computing, a fundamental departure from this tool model of software has happened. In particular, one of the fundamental changes was the emergence of online applications, i.e., networked computer-based systems where (1) multiple, unrelated users, (2) who do not own or control the server machines and/or most of the data being used, (3) provide personal data or assets as input to an online provider (4) to receive some output, in the form of information or not, delivered automatically or with human assistance. For example, in Google Web Search, (1) thousands and thousands of people, virtually simultaneously, (2) who do not own or control Google’s machines or data, (3) type everything from trivial questions to their most intimate desires (4) to receive a list of web-links as determined by a mostly-automated process which nevertheless include human experts (continuously trying to improve the search quality, protecting the service from malicious attacks from hackers, Google bombers, etc).

To better understand this transition, we have taken a theoretical approach where we try to understand such online applications with concepts, ideas, and analytical tools from services theory, which has been developed in the last 40 years mostly by researchers in business and management schools ([5, 9, 25] are good examples of textbooks in the area). We have found in our practice that the services theory framework is a powerful explanatory tool to understand many of the phenomena often observed in the design and development of online applications and, of course, notably the ones that are structured as services.

However, to correctly apply concepts from services theory to online applications, we have to carefully identify which applications have characteristics similar to traditional services. We start this process by discussing in section 2 that one of the main differences between online applications and personal tools is that in the latter the user controls the means of production: when and how intensively to use it, where the information (often personal) is stored and who can access it, and how much effort is put on a given task or goal. In other words, there is a high level of user dependency on the provider(s) of the means of production.

In the case of online applications, the factors of production belong to a service provider. As discussed later in the paper, this forces the online interface to deal with issues not usually even contemplated by a tool application: trust creation and maintenance; privacy concerns when handling sensitive information; communication of user context (often needed to satisfy a request); hard to predict interface response times due to fluctuations in demand; etc.. These are hardly the traditional requirements found in human-computer interface textbooks for an interface.

There is, of course, an extensive body of practice and empirical knowledge about developing interfaces for online interactive applications — exemplified by all the knowledge built in the last decade and half...
about web applications (for example, [11]). Also, there has been some work examining HCI and usability issues in e-commerce [10, 23], but we believe that these works suffer from not properly defining what an online service is and, therefore, face limitations when trying to create a bridge between online services and traditional services.

In fact, the core of our argument, presented in section 2, is that not all online interactive applications are service applications, but that the former can be divided along the dimension of how fundamental user input is for the production process, borrowing from recent theoretical work on distinguishing services from goods [17]. Based on the level of user input intensity, we draw a clear line in section 3 between online services providers (Google Web Search, Travelocity) and online information providers (cnn.com, Google finance). This distinction by construction defines online service applications as services systems, enabling “safe” application of knowledge, concepts, and methods of traditional services to online services.

We have been analyzing how traditional services knowledge can help the design, development, testing, and evaluation of online service applications. However, due to limitations in space, in this paper we address only human-computer interface issues to illustrate the approach. In section 4 we present some examples of how the services theory framework reveals more clearly where and how interface design and evaluation is affected by traditional service ideas. Although there has been some work examining traditional services theory concepts and how they apply in their counterparts in the online world — such as Ryan and Valverde’s study on waiting in line effects on consumer behavior [15], and Loiacono et al. work on service quality measurement of online services [8] — our approach here is to point out a much large number of candidate areas for future research. We finish this paper discussing issues related to the practical validation of this framework and examining future and possible developments of our ideas.

2. A Space of Information Applications

The core of our framework relies on some recent work in services theory by Sampson [16], popularized in [17], which proposes a new unifying definition for service processes. According to his work, a necessary and sufficient condition for a production process to be a service process is that “[…] the customer provides significant inputs into the production process.” [17, pg 331]. This primacy of customer input is put in contrast to manufacturing processes, where “groups of customers may contribute ideas to the design of the product, but individual customers’ only participation is to select and consume the output.” [16, pg 16]. Figure 1 shows two diagrams depicting graphically the main differences between manufacturing and service processes.

Notice that it is implicit in his definition (and discussed at length in [17]) that customers and service producer are separate entities. A better way to make this distinction is to say that the customer does not control most of the means of production. Our framework is based on taking those two key elements of a service production process, user input intensity and user dependency, and consider them as two dimensions in the space of information applications as depicted in Figure 2.

The first dimension relates to how much the user controls the means of production (the horizontal axis of Figure 2), or the level of user dependency on external providers. An application installed and able to run in a personal computer without network access such as traditional word processors (MS Word, LATEX) is a typical example of applications where the user controls most of the means of production. Other examples involve the basic core functions of the operational system, database applications using data stored in the user’s machine, personal back-up systems, etc. As the user loses more and more control of the means of production, the closer the application gets to typical web applications such as web search and browsing, information and news provision, web retail, multiplayer online games, etc. Typical middle-point applications are e-mail applications such as Eudora, Notes, or MS Outlook which combine local processing and storage with extent processing and data exchange with external servers; and local applications that depend on constant external updates such as most of today’s anti-virus programs.
The second dimension tries to characterize how much the user is an essential part of the input to the production process (the vertical axis of Figure 2), or the level of user input intensity, or simply, user intensity. As described by Sampson et al. [16, 17] the user can be the input to the production process in different forms: as herself (body or mind) such as when the services of a doctor in a hospital are seek; as her belongings, such as when a car is taken to a repair shop; or her information, as when giving financial information to get a loan from a bank. Notice that in all cases, the production process is unable to even start until the user provides her input. Typical information applications which require the user to be a significant part of the input to the production process are most of the interactive software we use in our everyday work such as word processors, spreadsheets, e-mail clients, computer games, etc.; and many of the web applications available today.

However, not all web applications require the user to be an essential part of the input to the production process. Typical cases are information or entertainment providers such as cnn.com, nytimes.com, or theonion.com. Although the delivery of particular pieces of information or entertainment is triggered by user input, a large part of the production process of the information is performed without any input from the user, through the manufacturing-like processes of news gathering and filtering, and entertainment production. Although the delivery of the information is interactive, the production of content is performed as free of user input as when cars are manufactured in an assembly line. Of course, nytimes.com is more dependent of user input than the The New York Times newspaper, but it clearly has a production process less dependent on user input than online services such as Google Web Search, Travelocity, or Amazon.

By taking these two dimensions spanning the space of information applications, we can draw the chart shown in Figure 2 that depicts different information applications as function of the level of user dependency (with the user being more dependent on external providers as we move from the left to right); and the level of user input intensity to the production process (with increasing user intensity from bottom to top). To illustrate our argument, we plotted some typical information applications on this chart in approximate positions. No metric for the two dimensions has been precisely defined, so the chart in Figure 2 should be regarded as a topological map showing the relative displacement of typical applications and services.

Finally, we are perfectly aware of the controversial aspects surrounding Sampson’s theory of services (see, for a summary, [17]), but the elegance of his framework is attractive and helpful to simplify most of our arguments in this paper. At the same time we believe our analysis is agnostic to the main contentious issues of his theory. Discussing in detail those issues is beyond the scope of this paper.

3. Online Service Applications

One way to understand the chart in Figure 2 is to consider the top half as the space of interactive applications; and the right half as the space of online applications. The top-right quadrant can then be seen as the one of online interactive applications which encompasses most interactive web applications. We observe here that not all the applications in this quadrant have significant similarity to traditional services, but only the top part of this quadrant where the user is a significant part of the input to the production process. As noticed before, traditional news and entertainment providers have a production process more similar to manufacturing than to services.

To take in account such issues, we propose the following definition. An online service application, or simply an online service, is an application where:

1. The user does not control most means of production.
2. The user (self, belongings, information) is a significant part of the input to the production process.
Having defined what we mean by an online service application, let us discuss typical characteristics of them. Figure 3 shows a typical architecture of an online service application. For example, a commonly found characteristic of an online service application is that the service provider has a multitude of customers, as shown diagrammatically in Figure 3. A service provider can exploit economies of scale to offer information provision at levels and cost that would be unfeasible in a tool-based approach. Imagine an extreme case where each of us had to run in our own machines a complete web-crawling system to be able to do web search. Instead, a service provider such as Google Web Search craws the web continuously, stores and indexes the data, and shares its results in the form of individual search results. All in exchange for some of the customer’s eyeball time — advertisement.

Having a large number of users to share the costs also enables online service providers to employ costly human domain experts and contact agents in hybrid production systems to accomplish tasks that are beyond today’s computing abilities, such as situations involving common sense reasoning or ethical analysis. Also it allows the shared use of human support agents to handle situations too atypical to merit the construction of a dedicated piece of software or interface, or when there is a need of human contact, for example, to evaluate how much to trust a customer complaint. This is hardly a possible solution in the context of tool software, where the user’s machine has to automatically solve every task. We think that such human-machine hybrid architectures, possible mostly in the context of online services, are a liberating idea for software engineering, traditionally submitted to the chains of full and complete automation.

Finally, economies of scale are just one of the advantages of online service applications. An online service provider can also take the part of a trusted and impartial intermediary between two or more customers, enabling environments suitable to the establishment of relationships. For instance, consider eBay and how it mediates buyers and sellers in its auction environment. Similarly, information about multiple users can be aggregated and used to establish “cast of thousands” data handling methods such as collaborative filtering [19] and social matching [21].

4. HCI of Online Service Applications

There has been very little theoretical work in terms of establishing a framework to understand online service applications and what is specific about how to
architect, design, engineer, evaluate, deploy, and manage them. Most of the discussion about the design and HCI issues of online service applications up to now tends to consider the broader class of online interactive applications (for instance, [11]) or the more restrictive class of online retailers [8, 10, 23].

By constraining our discussion to online services applications, we aim to be able to straightforwardly apply concepts and methods from traditional services theory. Notice that by construction, our category of online service applications as defined above assures that those applications are service processes, at least according to Sampson’s service theoretical framework [16]. Of course, there are many competing theories about how to characterize and classify services (see, for instance, the discussion in [17]) especially in comparison to manufacturing. It is quite beyond the scope and need of this paper to digress on the different views and to discuss how they may affect the HCI of online services. Instead, we assume from this point that our definition is sound, that is, that online service applications as defined in section 3 are services in the context of traditional services theory.

Although software production has always been considered by economists in the services category of businesses, a more careful analysis under the light of the discussion of the previous sections reveals that many of the production processes for software in fact resemble more manufacturing than services. This is especially true for shrink-wrap software, such as popular tools like Microsoft Office, Intuit Quicken, Adobe Photoshop, as well as basic and middleware software such as operating systems and database programs. Although there is user input during the design and implementation process, often from focus groups, individual user needs have hardly any impact on the software development process. Looking back into Figure 1, application and basic software have traditionally followed production processes that resemble more the manufacturing of physical goods than typical service production processes.

We believe that traditional HCI has thus been biased towards this model where tools are created for generic users to support a range of typical tasks. In terms of HCI evaluation and usability issues, a lot of effort is traditionally put in determining the typical usage scenarios of the tool and then to recreate in the laboratory meaningful test procedures.

The different dynamics of web applications required HCI practitioners to change their techniques to reflect the special needs of online applications (see, for example, the excellent manual by Nielsen [11]). Taking one step further, we believe that recognizing the specific characteristics of online service applications in contrast to generic online applications allows the development of more appropriate design and evaluation tools that also take in consideration the user dependency aspect. Moreover, we suggest using traditional services concepts and techniques as a springboard to facilitate the development of online services-specific HCI methodologies.

We structured our exploration work by considering the basic characteristics of services. We compiled and fused service characteristics listed by different authors [5, 9, 16, 25], arriving to a “compromise” list, which we believe most of them would agree to, consisting of: customer as input, heterogeneity, simultaneity, perishability, coproduction, and intangibility. We have looked into these six characteristics and examined how concepts and techniques from service theory related to each of them are likely to be relevant to the HCI design of online services. Due to space limitations, we present here some findings and ideas only for three of the six basic characteristics of services: customer as input, heterogeneity, and coproduction.

4.1 “Customer as Input” Issues

An immediate consequence of customers’ information as input in the context of an online service application is that trust, privacy, security, and authentication issues become key and strategic for the interface design. Unlike in traditional tool software where the privacy of data is often taken for granted and trust on the tool is often assumed to be unlimited, dealing with a service provider always involves an exchange of trust between the parties.

Customers often entrust online service providers with very sensitive information about themselves, their health, their finances, their loved ones, even their most intimate desires. The HCI research community has looked into issues related to trust in many different ways. As pointed by Wang and Emurian’s overview [24] most research suggests that trust in online applications is a function of “… a framework of trust-inducing interface design features, […] namely (1) graphic design, (2) structure design, (3) content design, and (4) social-cue design.” [24, pg. 21]. A study on web sites credibility by Fogg et al. [6] shares the same kind of recommendations, also present in a well-known set of design guidelines for online experiences by Shneiderman [20]. Featherman et al. [4] tested similar hypotheses and found them true in their experiments, although other factors seem to influence the perceived risk of an online e-payment service, including the computational literacy of the customer and the generic class of the online service.
When we look into traditional services knowledge and practice, the focus of techniques for building trust often focus on making the inner workings of services more “transparent” and visible to the customers. For example, a restaurant can improve the trust of their patrons in its cleanliness and service by having large windows to the kitchen area; or, in an online example, a shipping service may provide detailed real-time package tracking information. The difference is paramount: instead of asking for trust by improving the form of the interface, the service provider elicits trust by making its internal workings more visible: “trust what I do” instead of “trust what I say”. Interestingly, increasing the visibility of back-office operations in traditional services often improves also service quality, mostly due to the added pressure on the staff [7, pg. 160], but also by empowering the customer.

A services technique/tool that is often used to help identify the best candidates in the service process to be made visible to customers is a service blueprint. It is a map that portrays the service system, showing the whole process of service delivery, where customer contact happens, the roles of employees and customers, the visible elements of the service, and the overall flow of information (see [25, pgs. 267-276]). In particular, service blueprints depict clearly what is and is not visible to customers in a service process, by separating service components above and below a line called the line of visibility. We are currently experimenting with service blueprints in the design of online services to increase back-office visibility and, consequently, elicit more customer trust.

Dealing with the sensitivity of user information is also an issue that becomes fundamental in online services. Unlike in online information providers, private user information is often an essential part of the input to online services, for example, when applying for a bank loan. Traditional services often relied on the employees’ judgment to decide which information to ask a customer, which part to actually record, and to decide the trustworthiness of the information provided. Also, often the privacy guarantees were part of the human relationship between customer and employee. Unfortunately we are still trying to find ways to translate this human-based kind of privacy management to the online world. In the meantime, a general guideline is that, when an online service application asks for information that is particularly sensitive, the interface should clearly inform the customer why the system needs it, what the privacy policy is, for how long it will be kept, and whether there are alternatives are to provide that particular information. Marking clearly which elements of personal information are mandatory and which are optional is a good technique often employed by many websites. Notice that handling of private data issues are likely to be increasingly important given how common phishing attacks have become (see [3] for a nice study on why and how phishing works).

It is interesting that many websites still resort to the long, legalese-full license-agreement style of defining their privacy and data handling policies that are reminiscent of the never-read shrink-wrap software licenses. There have been efforts to simplify the establishment and negotiation between the customer and the service provider relationship, for instance in the SPARCLE project where privacy statements are machine-translated and automatically matched to customer-defined privacy standards [2].

4.2 Heterogeneity Issues

Users of software tools and customers of service software are, for all purposes, the same people. However, since people’s lives, needs, and desires differ substantially, the input to online services tend to be more heterogeneous than to other types of online applications because services involve a great amount of personal customer information as part of the input to the production process.

The effect of having highly heterogeneous input in the service production process is a traditional concern of services theory. First, heterogeneity of input requires the service designer to consider all possible instances of the input and how to handle all specific cases, including those instances where the service is not delivered successfully. It is interesting how traditional HCI research rarely tries to understand how to handle tasks which are not achievable, or even how to inform the user about the limitations of a tool. In contrast, service recovery, or how to handle unsuccessful delivery of services, is a major theme of research in services theory, given its known impact in service quality and customer loyalty. Zeithaml et al. [25, chapter 8] provides a good introduction to service recovery and to techniques used to alleviate the impact of failed delivery on customer satisfaction.

Another key issue that arises from customer input heterogeneity is ensuring consistency of service quality of the delivered services. It has been shown that heterogeneity of customer input, combined with the everyday fluctuations of the availability of human resources used in a service, create a vicious cycle that can drive service quality into a downwards spiral [13]. In this outstanding piece of research, Oliva and Sterman [13] also showed that among the key
elements to prevent erosion of service quality is the ability to monitor it adequately.

However, customers of services tend to have strong expectations about what they are entitled to receive from the service provider and how. An extensive body of literature in services theory has examined the role of expectations when measuring service quality (see, for a good discussion [18, chapter 2]). As a consequence, the most commonly used service quality instrument by the services industry, SERVQUAL [14], is in fact based on measuring the difference between service quality perceptions and expectations, or what is commonly known as the gap-model approach. There are many theoretical and statistical reasons to measure the gap between perception and expectation instead of simply determining the perceived quality of the service [18, chapter 2]. However, the most obvious advantage of using the gap-model approach is that it provides actionable information—which areas of the service are below what customers expect.

There is some strong evidence that, in fact, the gap model is also the right way to measure service quality in online services [22, 26], giving rise to a specific service quality instrument for web sites called WebQual [8]. Interestingly, user expectations and gap measurement have been used very sporadically by the HCI community (see [1] for a rare case). This work of validating and adapting SERVQUAL into the online-specific version, WebQual [8], is precisely the kind of approach we are advocating in this paper, applying well-established concepts from services theory into online service applications.

We are also starting to believe that this heterogeneity of input and output questions the very core foundations of the HCI practice. In particular, we have seen how difficult it is to perform traditional user-centered design in the context of the lack of prototypical users and tasks created by heterogeneity. Not only it is extremely difficult to cover a reasonable spectrum of customers during usability tests, but also it is hard to recreate in a laboratory the right context, diversity of tasks, and expectations. This is corroborated by the often common practice of web developers of circling around the heterogeneity issue by using extremely fast prototyping methods so beta versions of the service are not tested in a laboratory but directly with a large number of actual customers.

### 4.3. Coproduction Issues

Software developers are traditionally trained in a mindset where the goal of software is to automate a task, given some input from the user. In that framework, it is not a surprise that the user is seen as an outsider of the production process. Notice that the word “interface” itself expresses an idea that the user is external to the system. Contrarily, in traditional services integrating the customer to the production of the service — or coproduction — is a well-known technique to decrease labor costs and increase customer satisfaction (see [25, chapter 13]). Further, coproduction often is used as a way to empower their customers and not simply to exploit them.

Coproduction has often been used in online applications, though often disguised and many times misunderstood. Google Web Search is based on the notion that the customer can do a lot of information filtering himself as long as a reasonable summary is provided and the response time is fast. Similarly, online travel service providers such as Travelocity have pushed most of the travel agent’s job to the customer. In many of these cases there is some loss of quality when part of the production process is moved to the customer, since less expert knowledge is brought to bear by human professionals. However, we should also recognize some key benefits of coproduction in those cases. For instance, direct access to information about travel gives the customer more time to reflect and weight options without the pressure of making a decision. As much as not having a waiter ready to clean up the table at a fast food restaurant allows customers to prolong their stay as they wish, coproduction in an online service application can be used to break down different steps of the production process in a pace that is more convenient and pleasant for the customer.

However, HCI practitioners should be careful about how coproduction works and its impact in their evaluation techniques. For instance, the duration of a task, often used in usability studies, is not an appropriate measure when customers are taking time to decide among different options, gathering more information, or weighting risks. Also, coproduction often involves some level of customer training, so interface design has to consider carefully how the customer is going to learn the skills needed to coproduce effectively. There are many interesting teaching techniques that have been developed by traditional services (see [25] for some examples) and the HCI community can definitely learn from them.

### 5. Discussion and Future Work

At this point the reader can be understandably questioning where the evidence is that the service theory framework and techniques are really useful for online service applications. There are at least two documented cases — the line waiting study of Ryan and Valverde [15] and SERVQUAL/WebQual case of...
Loiacono et al. [8] — where traditional services techniques are shown to be more appropriate to the online services domain than traditional HCI methods.

We have also shown here, notably in the previous section, many examples where we could suggest methods and techniques from traditional services that address common difficulties facing online services interface design and evaluation. But most importantly, we expect this introductory discussion about how services theory concepts impact online services interface design to create questioning and curiosity in the field and trigger further research. In our particular case, we are focusing efforts in codifying and adapting traditional services concepts and methodologies to the design and evaluation of online service applications. For instance, we have been trying to use service design [25, chapter 9] methodologies to reinvent the user experience in online technical support.

From online multiplayer games to web search, many of the new components of the online information landscape are structured as services, but they are still often designed and evaluated under a “tool” view of computing. We hope we are contributing here to change this mindset, towards increasing not only the quality of customer experience but also the efficiency and adequacy of online service applications.

References


