Cross-Site Personalization: Assisting Users In Addressing Information Needs That Span Independently Hosted Websites

Kevin Koidl
Trinity College Dublin, Ireland
College Green, Dublin
Kevin.Koidl@scss.tcd.ie

Owen Conlan
Trinity College Dublin, Ireland
College Green, Dublin
Owen.Conlan@scss.tcd.ie

Vincent Wade
Trinity College Dublin, Ireland
College Green, Dublin
Vincent.Wade@scss.tcd.ie

ABSTRACT

This paper discusses Cross-Site Personalization (CSP) an approach to provide personalized assistance for the user in addressing information needs that span independently hosted websites. This is done by seamlessly personalizing the support offered to each individual user, as they browse across multiple websites, by modeling the user's interactions and then augmenting information access points, such as links, on each independent website. Cross-Site Personalization is realized as a third-party API offering Personalisation As a Service to ensure cross-site and cross-device usage. The personalized augmentations are provided through module extensions for the Web-based Content Management Systems (WCMS) Drupal. The approach is non-intrusive and does not limit or alter the user's information access paradigm. This is done by visually augmenting the existing hyperlinks on webpages. The design of the API ensures user's privacy by not disclosing personal browsing information to the websites. Rather, this approach recommends how each website may adapt their information and navigation structures to meet user's information needs. Finally, the approach ensures user control and scrutiny. The user can enable/disable CSP at any time and view any information collected. The evaluation of the approach was conducted with a real-world use case. This paper introduces the architecture, a prototype implementation and encouraging evaluation results.

Keywords

Web Personalization, Information Access, Hyperlink Augmentation, Adaptive Hypermedia, User Privacy, User Model Scrutiny and Control, User Experience, Adaptive Linking, Networked User Engagement.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

HT'14, September 1–4, 2014, Santiago, Chile.

Copyright © 2014 ACM 978-1-4503-2954-5/14/09...\$15.00.

<http://dx.doi.org/10.1145/2631775.2631798>

1. INTRODUCTION

The Web has become an essential commodity within modern societies allowing the access, creation, management and distribution of vast amounts of information. This has led to an information explosion and has resulted in an increased difficulty to organize digital information based on the user's current and constantly changing information needs. To allow users to find and use information most effectively, it is essential to develop novel and innovative personalised assistance approaches helping the user find and explore relevant information. Several publications across different research areas, such as library and information science [1][2], cognitive science [3], social science [4] and medical science [5] have voiced this need, and the failure to address it properly can result in increased costs to the economy and the wider society [6].

In recent years, computer science has addressed this challenge by introducing systems that assist the user's information needs by applying a user model reflecting the users interests. Examples of systems using an interest focused user model are keyword-based search engines [7][8][14], browser based support mechanisms, such as bookmarking and browsing history tools [9][10], recommender systems to identify and recommend content related to the user's current interest [11] and Web personalization techniques that adapt different aspects of the web experience to the needs and preferences of the individual user and based on a individual user model [12][13]. In addition collaborative approaches were introduced. The main difference is that collaborative approaches implement a '*People that like x also liked y*' approach. One example is Amazon's product recommendations service [25][50]. A second example is Facebook's '*socially connect*' to recommend content that friends and peers 'like' [15].

Even though current approaches are successful in providing fast and accurate results for simple information needs, such as finding information related to a product via personalised search or collaborative recommendation system, most approaches cannot assist the user in more complex needs. For example, if a user's information needs span different subject domains, the user may need to gather information from several independently hosted websites. This cross-site browsing process can introduce friction based on different website layouts the user has to adapt to in order to identify the relevancy of the navigation choices presented [29][42]. To reduce this friction a browsing support approach is required that can unify the mostly fractured browsing experience across independent and unrelated websites [39]. Moreover users may choose to use multiple browsers and devices to address more

complex information needs. Therefore, any supporting approach has to ensure that the user remains free in her choice of information access (not limiting the browsing space) and her choice in browser/device; thus not limiting the browsing process to a specific browser based plugin and/or device.

This paper introduces an overarching, service-based approach towards assisting the user in information needs that span independently hosted websites.

This paper is organized as follows. This introduction is followed by a discussion related to background and related work. After this design considerations are introduced that allow the addressing of the discussed challenges. This is followed by a brief discussion of the implementation. Finally, evaluation results are discussed.

2. BACKGROUND AND RELATED WORK

Providing personalized browsing support across independent websites is difficult due to the lack of communication between the websites. This can lead to a browsing experience that is mostly confined within a website; thus if a user is gathering information about a specific topic on one website and then decides to browse to a different website, the latter will not be informed about the user's previous information needs [16][17]. Furthermore any personalization techniques used by one websites will not be able to assist the user in relation a wider information needs across a second (independent) website. The effect of this siloing of information and personalization techniques can lead to the repetition of user actions, such as keyword based search queries, and navigational patterns throughout the different websites, such repetitive interaction is both frustrating and time consuming for the user [18].

A key challenge therefore remains in providing Web personalization techniques that assist users across the web and not only on isolated Websites. To address this problem Web personalization techniques have to balance both the needs of the website user and website publishers. For website users the introduced Web personalization technique should provide assistance across different websites, but in doing so honor the user's privacy needs and browsing freedom. For the website publisher the introduced Web personalization techniques should include simple and cost effective integration of Web personalization to existing websites and honor the website owner's control over the website. For example a publisher might want to ensure specific content is promoted based on commercial goals even if it the content does not reflect the current interest of the website user. Motivated by the discussion above following Figure 1 illustrates a high-level abstraction of the problem space a Cross-Site Personalisation (CSP) approach needs to addresses.

Figure 1 Cross-Site Personalisation Abstraction

Based on this illustration (Figure 1) it can be argued that a CSP approach requires influencing both areas (needs of the website user and website publishers) simultaneous and in real-time by creating a *state of equilibrium* in which the needs of the user and the website publisher are balanced. This state is introduced as *Equilibrium of Web Personalisation*. To ensure the needs of both sides are addressed adequately this paper argues that the resulting CSP assistance should be non-intrusive [37][38][40] for the user (not limiting the users free browsing paradigm) and non-invasive on the website (ensuring simple, light weight integration).

Within non-intrusive personalisation approaches three techniques stand out: Information Retrieval [51][22], Recommender Systems [50] and Rule Engine based techniques [52]. All three techniques have proven successful in personalization use cases, with limited impact on the users free browsing paradigm.

However, these techniques are not distinct, for example Content-based Recommender Systems are often based on Information Retrieval techniques [26][53]. Furthermore, many non-intrusive approaches have been extended towards personalisation since they have been introduced [54]. However, the main shortcoming of non-intrusive approaches still is the lacking focus on specific user needs and preferences. Most approaches apply a more generalized data driven approach resulting a 'User that viewed this page also viewed following pages' personalization scenario.

In relation to introducing CSP to websites in a non-invasive manner, two approaches can be discussed: (1) Non-Invasive integration of web personalisation techniques at design-time and (2) non-invasive integration of web personalisation at run-time.

Non-invasive integration at design-time refers the research and application field of Web Engineering, which addresses the shortcoming of integrating techniques and methods to existing websites. Here, a structured modularization of the website development process is introduced [55]. This process relies on modeling the overall website based on software engineering principals. Several web engineering design methods were introduced. The main design methods introduced are OOHDM [56], UWE [57] and OOWS [58]. In addition data modeling approaches, such as WebML [59] and WebRatio [60] were introduced as model-driven engineering environments.

The advantage of Web Engineering is in relying on strict design methodologies. For example OOHDM (Object Oriented Hypermedia Design Method) has following phases: requirements gathering, conceptual design, navigational design, abstract interface design and implementation. The result is UML like diagrams and re-usable design pattern. Most Web Engineering approaches include functionalities that assist the user in finding and exploring information. This includes techniques discussed above, such as Information Retrieval and Recommendation Systems [54].

The main shortcoming of Web Engineering is lack in flexibility in extending the websites functionality beyond what has been stated in the websites setup design. Furthermore any extension towards CSP would have to be considered in the early conceptual phase with limited flexibility during run-time.

In contrast the integration of personalisation methods and techniques at run-time is even more challenging. The main reason is the missing understanding of the interdependencies within a website. A promising approach is the use of website frameworks. Such frameworks are often referred to as Web Information Systems (WIS) and allow the implementation of an entire website

as out-of-the-box deployment [61][62]. In certain environments this approach enables non-invasive extensibility during run-time.

However, WIS are traditionally proprietary software products limiting the extensibility without the vendors help and/or approval. A promising solution to this limitation are Web-based Content Management Systems (WCMS). They allow a simple and more flexible implementation, based on their Open-Source nature. The most important feature of WCMS is the extendable framework, which allows external modules to influence different level of the functional layer of the website without the need of re-design or re-deployment. The main shortcomings of WCMS are a lack of strict and coherent coding guidelines and documentations.

Applying CSP to WCMS provides the added benefit of giving the website owner control in deciding what area of the website should be affected by the link guidance provided by the CSP service.

A further related research area is known as Adaptive Hypermedia (AH). AH traditionally focuses on problems known as *information explosion* [27][28] and its resulting phenomena known as *information overload* [30][31] and *lost in hyperspace* [32]. The former refers to the user being overwhelmed by the amount of information choices. The latter refers to orientation difficulties of the user within and across different websites.

AH approaches seek to assist the user in addressing such phenomena by tailoring the web experience to the specific needs and preferences of the user (or a set of users) and at the same time avoid that the user is limited in its information freedom [19][20]. This is done by providing the information users want or need without expecting them to ask for it explicitly [21][23]. Methods and techniques enabling Web personalization have become important for both research and commercial applications [21][24]. Especially in domain specific application, such as e-Learning, the use of AH techniques and methods have been proven to be successful by increasing the learning performance of the student [33][35].

This domain focus however has also lead to a shortcoming known as the open-corpus problem described as: *'The problem to provide adaptation within a set of documents that is not known at design time and, moreover, can constantly change and expand'* [34][36].

This paper argues that the open-corpus problem of AH needs to be extended towards the *open-web problem* of AH for it to reflect personalization on the open web by adding *'[...] and which may reside in individually hosted websites on the open web.'*

To address the open-web problem of personalization this paper discusses the concept of Cross-Site Personalization (CSP) as a process in which a web user is individually assisted in addressing information needs that span independently hosted websites.

Finally, in relation to web personalization privacy and trust related issues have been discussed extensively [41][63]. A major shortcoming of AH and web personalization in general is the danger of filtering the information in a way that is not in the interest of the user; thus limiting the users information space. A typical solution is to provide transparency. This has been introduced by [40] as user model scrutiny. Here, the main limitation is related to the visualization of the model. The discussion of information visualisation however is out of scope for this paper. For an in-depth discussion about model visualization strategies see [40] and [64].

The following section introduces the architecture of the proposed CSP approach.

3. CSP ARCHITECTURE OVERVIEW

This architecture provides the conceptual basis for an introduced prototype implementation. The goal of the introduced architectural overview is to illustrate the responsibilities and capabilities of the different architectural components. In the following high-level architecture (Figure 2) is proposed and subsequently discussed.

Figure 2 High-Level CSP architecture

3.1 Term Identification Component

The term identification component of the CSP architecture has the responsibility to facilitate the identification of terms. The terms indicate the meaning of the underlying content from the websites within the cross-site browsing space. This identification can be a result of using already existing (editorial) terms provided by the website or by using additional Natural Language Processing services/tools to identify terms (e.g. OpenCalais¹). The main responsibility of this component is to ensure the creation of a *shared conceptualization* of the user's CSP enabled browsing space. This shared conceptualization is represented as term space and is based on the terms extracted from the content of the websites the user is browsing across. Once the terms are identified they can be used by the user profile component to represent the user's information needs. Furthermore, the term identification component requires means to facilitate continuous content updates to ensure the shared conceptualization of the user's cross-site browsing space is up-to-date. For this the term identification component needs access to the content on the different websites. Within the interface layer of this architecture it is proposed that the website sends the content to the website that is to be used for CSP. This ensures the independence of the approach from the websites.

3.2 User Profile Component

The user profile component is responsible for the correct storage and aggregation of terms related to the content the user is browsing. These terms are provided by the term identification component and are stored together with the engagement activity the user is conducting on the content. This activity can include

¹ <http://www.opencalais.com/>

mouse clicks, scrolls and keyboard presses. Together the terms and the activity form the base to create an informed decision by the strategy component. Furthermore, the terms are used within the Scrutiny Component to enable the representation of the user's profile. The User Profile Component requires the identification of the user. This identification can either be provided directly by the website or by a sign-on mechanism, such as OAuth.

3.3 Strategy Component

The strategy component is responsible to provide an *informed decision* to the website the user is currently browsing. An informed decision assists a website to identify relevant links based on the users overall cross-site browsing. It implements a decision process that facilitates elements of the user's current user profile and relates this with the current website the user is browsing. The resulting *informed decision* allows a website to receive relevant information to provide adaptive guidance based on the user's overall information needs. The component implements techniques that ensure fast and flexible calculation of an informed decision. The calculation has to take the users recent and long-term user engagement into account and also takes a depreciation factor is taken into account. This allows long terms interests to be reflected but not as strong as recent interests. Finally, the strategy component has to ensure that the user's privacy is honored. This is achieved by not passing any information related to the users cross-site needs to individual website. The design approach followed here is to allow the service to provide information on the relevancy of links, but to not provide information on why the links are relevant. The informed decision is closely linked with website extension. These extensions allow the simple application of CSP based on the informed decision.

3.4 Scrutiny Component

The scrutiny component addresses requirements related to privacy, trust and control needs of the user. It allows users to access their user profile and to influence the profile. Furthermore the user can receive information on where and how the user related data was collected and used. In its basic implementation the scrutiny component allows the visualisation of the terms that correspond to the webpage the user has accessed. Furthermore size of the terms can indicate the relevancy of the terms based on the users engagement with the different websites. This is visualized as term cloud that can be only accessed by the user.

3.5 Interface Layer

The interface layer provides an abstraction from the specific implementation of the different websites within the user's cross-site browsing space. For this it implements a RESTful API (Advanced Programming Interface) to facilitate the communication between the CSP Service and the websites the user is browsing. A key responsibility of the API is to ensure that the interface with the CSP approach is device independent and does not depend on a specific technology stack. Furthermore, it ensures fast and accurate interconnectivity between the interfacing websites and the CSP Service. The API can either interface with a website through bespoke integration or through provided website extensions.

3.6 Web-based Content Management System Module Extensions

Web-based Content Management Systems (WCMS) module² extensions allow a simple and non-invasive integration of CSP to existing website implementations. The responsibility of the WCMS module extensions is twofold: (1) To facilitate the communication between the website and the API of the CSP Service and (2) to provide non-intrusive personalization's to the user, within the website the user is currently browsing, based on an informed decision from the CSP Service. By using WCMS extensions it can be assured that CSP provides non-intrusive assistance for the user and non-invasive integration in existing websites (during run-time).

3.7 Cross-Site Browsing Space

The application of CSP to independently hosted websites introduces a Cross-Site Personalization Browsing Space. Within the browsing space the user receives CSP through navigational assistance. The navigational assistance is enabled either through direct integration with the API of the CSP service or by using the above introduces website extensions. Based on the back-end integration of the CSP service with the website it is possible that the user decides when the CSP is enabled and when disabled. This allows the user to control the mechanism and ensures the user's privacy is honored. The enabling/disabling of the service can either be facilitated through the user signing into the website or by directly signing into the service.

The following section discusses implementation considerations of the above-introduced CSP architecture.

4. IMPLEMENTATION

This section describes the technological architecture required for CSP based on the architectural overview discussed in the previous section. Figure 2 has been extended with implementation details (see Figure 3). In the following the implementation decisions of the different components are discussed.

Figure 3 Implementation Overview of CSP

² For this research the WCMS Drupal (<https://drupal.org/>) was used.

4.1 Term Identification Component

The term identification component enables term identification as a constant and fast process. Furthermore it allows the identification of terms across several different subject domains, possibly including different languages. Based on this an extendible architecture was implemented. This allows the adding of additional external term identification services. Due to its reliability the current iteration of the CSP approach interfaces with OpenCalais³. An addition investigation was conducted in implementing a solution that provides the same flexibility and maturity as an external term identification service, but is deployable within the CSP Service. OpenNLP⁴ was identified as one possible solution. However, it requires an extensive training phase and relies on pre-set (domain specific) content to train the service. Based on this shortcoming the Apache Stanbol Project⁵ was investigated. It implements a pre-trained OpenNLP instance. However, at the time this research was conducted the project was still in incubation phase.

The process of term identification is initialized as soon as a website enables CSP (based on this topic by deploying an enabling the CSP extension module). After the website publisher enables the CSP module the module sends the content body, title, path id and any associated terms of the openly accessible content to the API of the CSP Service. The service then requests the term identification from OpenCalais for the content body sent, which returns social tags associated with a relevancy of either 0.5 or 1. The CSP Service stores the identified terms and the identified relevancy together with the path and website id. The CSP Service therefore implements a table representing all terms identified from the content of the CSP enabled website.

4.2 User Profile Component

The user profile is stored in MySQL and stores the terms related to the user's page views and related browsing activities on the page. A user activity is represented by a unique user id, a unique website id, the content path within the website, a list of the inferred terms related to the content path, scroll counter, click counter, touches counter, cut-and-paste counter, key pressed counter and timestamp. The timestamp is important to indicate when the last page access has taken place. This allows the strategy component to ensure more up-to-date activities are prioritized.

For user identification OAuth⁶ was implemented. This mechanism allows the website to remain open by allowing the user to sign-in to the CSP Service via the website. Furthermore the approach enables users to authorize the website to receive an informed decision to assist the user.

4.3 Strategy Component

The strategy component implements an information retrieval query generation approach. It is implemented based on Apache Solr⁷ and uses the existing content index built by the Term Identification Service. Once a website requests an informed decision the strategy component requests the user's current profile terms and the calculated relevancy of the terms. The terms and the related weight are used to create a query on the stored index of the

website that requested the informed decision. The result list is filtered based on the links the website has requested an informed decision for. The result of the filtering is a list containing links and the calculated relevancy (based on the relative position within the result list). The usage of IR was chosen due to its flexibility in relation to term usage and its performance.

Finally, in relation to data privacy requirements it is important to note that no terms are sent to the different websites. The website only receives a list of links with related relevancy indications.

4.4 Scrutiny Component

The scrutiny component allows the user to see the terms in the user profile together with the indicated relevancy as size. For this a word cloud can be used.

4.5 WCMS Client Integration

To enable a website to offer CSP to their users, a custom build Drupal 7 module extension was deployed.⁸ The role of the module is:

1. To enable initial term identification within the CSP Service by sending content to the CSP Service.
2. To provide a sign-in mechanism for users to enable CSP. This sign-in is provided on the website the user wants to enable CSP on and is implemented via the CMS plugin. An additional sign-in is provided on the service website where the user can see and manage their user profile.
3. To enable user behavior tracking once the user has signed in.
4. To provide non-invasive visual assistance once the user is signed in and has authorized the website to receive an informed decision for CSP.
5. To send content updates to the CSP Service on a regular base independent from user being signed in or not.
6. Request an informed decision from the CSP Service.

Based on the API approach taken by the CSP Service the deployment of extensible modules is not necessary i.e. the website publisher and or the wider open-source community can chose to write addition website extensions/plugin-ins. However, using the provided extension significantly decreases integration time and based on the extensive research in relation to non-intrusive visual assistance ensures that the user need in not being aversively guided or controlled is addressed.

The WCMS module extension is focused on a menu structure that is introduced as separate element or replacing element on a Drupal powered website. The website publisher therefore can replace the standard menu element with the CSP enabled menu element. The menu area looks identical and only is affected if the user enables CSP. The menu structure is tied in with the request of an informed decision. In this use case implementation the influence of the CSP Service is confined to the CSP enabled menu. Therefore the informed decision influences only links within the CSP enabled menu. Prior to rendering the webpage after the user has clicked the CSP module sends the links that the user will see on the resulting webpage to the CSP Service. The CSP Service generates an informed decision that is rendered within the CSP enabled menu.

³ <http://www.opencalais.com/>

⁴ <http://opennlp.apache.org/>

⁵ <http://stanbol.apache.org/>

⁶ <http://oauth.net/>

⁷ <http://lucene.apache.org/solr/>

⁸ Drupal was used in this use case implementation based on its wide adoption.

5. EVALUATION

This evaluation section introduces the result of a use case based user study in the tourism sector.

5.1 Experimental Setup

This experiment focused on a use case in the tourism space. The content was provided by Failte Ireland and is used live⁹. The experiment consisted of three websites. The first website acted as entry website and was based on health and fitness related content in the four activities Golfing, Hiking, Walking and Cycling. The purpose of the entry website was to train the users profile to enable CSP across the remaining two websites. The two additional websites were related to tourism and were deployed each relating to different locations (Killarney and Dingle). Each website was hosted as a separate WCMS implementation with different templates to ensure a different look and feel on each website. To evaluate the effect of CSP on the user's performance and to assess the self-reported feedback it was conducted as a between-subject experiment, with each cohort consisting of equal numbers. Each participant was randomly assigned to the CSP enabled or CSP disabled group. Two different system settings were used. The first system setting consisted of all four websites with CSP enabled the second with all four websites CSP disabled. The underlying websites including the layout and content were equal in both system settings.

Overall 60 participants took part in the user-focused study. The participants were sourced through school mailing lists and black board notices throughout Trinity College Dublin. The experiment was conducted in a closed lab environment to reduce distraction, and allowing further feedback through an unstructured interview. The evaluation data was collected through questionnaires, system logging and unstructured interviews.

The duration of each evaluation was approximately one hour and the overall evaluation including the two-phased design study was conducted over a timeframe of six months. One of the goals of the evaluation was to ensure the experiment setup was as natural as possible. To ensure a realistic browsing environment the tasks were based on real-world fact gathering tasks and the experiment environment was set up without any intrusive recording devices.

Each participant was asked to conduct one browsing tasks related to planning a holiday around the activity of Golf. The goal of the experiment was to explore and collect information that would allow the participant to convince friends that a holiday around the activity Golf is worth pursuing. After filling out the ethics form and informing the participant what information will be recorded, a short description of the task was presented. The participant was asked to provide all answers in handwritten form to ensure the participants were able to answer the task questions without interrupting the browsing setup (which would have been necessary in an online form). Finally, the participant was given a chance to ask questions. Once the participant was ready to start, task time was recorded and the examiner left the room. The participant was asked to stop the clock once the task was completed and then call the examiner back into the room.

After the task the participants were asked to complete an online survey with 22 questions. The cohort with the personalised setting was asked to fill out a second short paper questionnaire with the examiner in the room in order to trigger open questions and

responses in relation to the introduced CSP approach. For this the participant was made aware of the introduced approach.

5.1.1 Baseline Comparison

The experiment was conducted comparatively as a between-subject design. Each cohort was presented with a total of three independently hosted websites. The first website was used to train the participant's profile. The latter two evaluated the appropriateness of the CSP approach. All three websites were hosting real world content in the area of recreation and tourism. The baseline system was deployed without any personalization features enabled. From the participants perspective there was a difference how the side menu presented links. The CSP enabled websites implemented a green link annotation icon beside every link in the menu. Each annotation icon was filled depending on the level of relevance the link had (based on a relevancy scale from 1-10). All other website features were identical. The following discussion of the findings is split into user performance and self-reported feedback. Note following image (Figure 4) for an overview of the annotations used.

Figure 4 Open domain use case example

The annotations were implemented as green icons in the menu structure on the left hand side. This annotation was consistent across all websites ensuring the user was able to become familiar with the approach. It also has to be noted that only annotations were used. This forms an overlay on the existing visualisation of the page without applying content filtering or sorting. The icon was chosen based on user trials with different shapes and relevancy indications.

5.1.2 Results User Performance

The findings related to the user's performance is divided into click-related and time-related findings.

Click-related Findings

To discuss the influence of the CSP approach on the click behavior of the user, the overall clicks were analyzed.

The comparison of the total clicks (without outliers and including repeated page views) within the cohort assessing the CSP enabled websites was 173 vs. 268 in the CSP disabled websites. The individual results were 83 (enabled) vs. 111 (disabled) clicks in the first 'Health and Fitness' website, 52 (enabled) vs. 99 (disabled) in the Killarney website and 38 (enabled) vs. 58 clicks (disabled) in the final website 'Dingle'.

⁹ <http://www.discoverireland.ie>

To further investigate this encouraging result the average clicks of each individual website were compared.

The highest difference in the mean value was in the ‘Killarney’ website. To identify if the difference in the mean values are statistically significant a t-test assuming equal variances with a hypothesized mean difference of 0 was conducted. The corresponding p-value resulted in 0.169 for ‘Health and Fitness’, 0.003 for ‘Killarney’ and 0.448 for Dingle concluding that there is a significant difference between the mean click values of ‘Killarney’ in both cohorts. A reason for this difference may lie in the fact that each website presented a different amount of links the participant could choose. On the ‘Health and Fitness’ website all five links were relevant. The Killarney website participants were presented with 14 links of which 4 were relevant. In the Dingle website 8 links were displayed of which 2 were relevant. The significant difference in the decrease of clicks in the crossover from the first to the second website can be judged as evidence as a positive impact of CSP on the users performance.

Finally, cumulative data was analyzed. Cumulative observations were not directly comparable with the observations above due to using all three website datasets as one dataset. Therefore the following comparison only serves investigative purposes. Overall the average values resulted in 10.42 clicks (median 9) for the CSP enabled websites and 14.3 clicks (median 12) in the CSP disabled websites.

Task Completion Time

The overall task completion time (in seconds) within the CSP enabled cohort was 23100 and within the CSP disabled cohort 21960. The overall mean times resulted in 1073.8 seconds vs. 933.6 seconds.

Following image (Figure 5) illustrates the overall mean time across the different website.

Figure 5 Overall Mean Times CSP in Open Domains

Interestingly, the gap between the means decreases. This trend is important and possibly indicates a performance increase of the introduced CSP approach as the users continue browsing across separate websites. Most encouraging is the decrease within the last website.

5.1.3 Results Self-Reported Feedback

To gather the self-reported feedback, the participants were asked to provide answers to 26 questions. The questionnaire had the objective of assessing the perceived appropriateness of the CSP approach, and whether it assisted users in addressing information needs in open domains. For this the participants were asked direct and indirect questions. Direct questions addressed key aspects of the CSP approach (e.g. perceived difficulty in browsing) and

indirect questions assessed cognitive indicators, such as the perceived motivation and frustration.

The questionnaire was presented on a custom made online interface. The participants were presented with a slider to provide the answer on a scale from 0-100. Each participant was presented with visual markers at 0 for strongly disagree, 25 for disagree, 50 for neutral, 75 for agree and 100 for strongly agree. Yes and no questions were binary (0 for yes and 100 for no).

To assess the overall perceived mood of the participants both the perceived effect of the CSP approach on frustration and motivation was queried. The result indicated no statistically significant difference between both cohorts although with the average being slightly higher for the participants assessing the CSP enabled websites (averages 15.28 vs. 19.46 and median 15 vs. 11). The low level of frustration coincided with a high-level of motivation with the average of the CSP enabled websites feedback slightly higher (averages 80.29 vs. 74.93 and median 85 vs. 76).

To assess if the task was appropriate the participants were asked if the task reflected a real-world cross-site browsing task. Most users agreed with similar means in both cohorts (averages 75.52 vs. 76.65 and median 76 vs. 77). This result is encouraging and indicates that the participants were not negatively biased due to the pre-set nature of the task and the unnatural browsing setting with a lab.

To assess the appropriateness of the CSP approach participants were asked to rate the perceived influence of system guidance. As mentioned in the experiment setup, no participant was informed about the CSP approach prior to the experiment. The result is therefore encouraging (average 79.55 vs. 59.41 and median 78 vs. 58) with the mean being significantly higher than the CSP disabled (p-value of 0.0006). This result can be interpreted, as a key finding in relation to the participant’s perceived appropriateness of the approach to assist in cross-site information needs.

A further encouraging result was found in the question related to the difficulty in browsing. The answers result in a significant difference towards the CSP enabled system setting (p-value 0.0216) with an average of 24.07 (media 25) in the CSP enabled based feedback and 36.52 in the CSP disabled 36.52 (median 38).

Finally, an interesting result was provided by the participants in the question related to privacy concerns. Here, the cohort assessing the CSP enabled websites indicated a higher average concern than the cohort assessing the CSP disabled websites (59 vs. 55.79 and mean 46 vs. 48). This is surprising given the fact that this question does not directly relate to CSP or the usage of personalization in general. In relation to this research the result can be interpreted as a higher awareness for the CSP approach within the cohort assessing the CSP enabled websites. Furthermore some participants indicated in the open interview at the end of the experiment that they felt watched and/or guided – a perception that might make users more curious about the source or motivation of the approach. Some participants stated they started to trust the guidance after a while and that they look for the visual indication first when accessing a website.

6. CONCLUSION AND FUTURE WORK

This paper discussed Cross-Site Personalization novel approach to assist users in addressing information needs that span independently hosted websites.

The Cross-Site Personalization approach was introduced as third-party API offering Personalisation As a Service. The personalized link augmentations were realized through module extensions for the Web-based Content Management Systems (WCMS) Drupal. The introduced architecture ensures not limiting or altering the users free browsing paradigm. Moreover, the design of the API ensures user's privacy by not disclosing personal browsing information to the websites. Finally, the approach introduced ensures user control and scrutiny by allowing the user to enable/disable the CSP assistance at any time and view any information collected by the service.

A prototype implementation was presented and subsequently evaluated. The evaluation indicated the usefulness and appropriateness of CSP a real-world use case. The results were encouraging motivating following future work.

Additional Use Case Evaluations

To ensure the above-introduced CSP approach can be applied widely on the open web further real-world CSP use cases need to be investigated. The main challenge in identifying such use cases is to ensure that they are motivate by complex information needs that require the user to browse independently hosted websites. These use-cases are among others retail (gathering reviews and feedback related to purchasing a products), e-Learning (gathering information to gain knowledge e.g. informal learning or conducting research projects) and Online Customer Care (gathering information related to solving a problem).

In addition the evaluation of CSP use cases needs further investigation. The above-introduced evaluation was conducted in a controlled lab setting. The main motivation for this was to allow unstructured interviewing and strict measurement of metrics without any distracting or disturbing factors. However, it can be argued that this artificial environment may limit the results. A more natural and familiar environment however would require a much large user group and sufficient A/B testing, which can be both time and resource intense. Moreover due to the different interests and interests of users (which would be uncontrolled in an open scenario) it would be unclear if the user clicked on a link due to a general interest (i.e. would have clicked in any case) or due to the personalized augmentation (i.e. is only clicking because of the augmentation).

On the contrary ensuring full insight the experiment would require the usage of more intrusive observation methods, such as eye tracking and/or video surveillance [44]. However, intrusive observations interfere with the natural open-web browsing behavior of the user. Therefore a balance between more intrusive observation and the participants' natural browsing environment should be found and aligned with additional use cases (possibly using browser plugins tracking users over a prolonged period). These evaluation challenges are not new to the research field of web personalization and have been highlighted throughout different publications [45][47].

Networked User Engagement

Networked user engagement is becoming increasingly popular with larger corporations owning multiple web properties. The most prominent examples are yahoo and AOL [48]. For owners of larger website federations it is becoming increasingly important to both understand and assist the user across the commonly owned web space. Based on the cross-site notion discussed and applied in this paper, future work can be conducted by using large datasets from a larger interconnected set of websites.

User Profile Management and Social Media

The CSP approach presented in this paper allows the user to view their interest model (model scrutiny) that relates to their overall cross-site information needs [49]. Through comments, several participants expressed their interest in actively engaging with their user profile and sharing it with peers and friends.

Flexible Intrusiveness

A core element of this research is to ensure that the introduced approach ensures non-intrusive assistance for the user and non-invasive integration to the website to address website publisher needs. The commonality of both aspects (non-intrusive assistance and non-invasive integration) is to balance the needs and requirement of the user and the website publisher (state of equilibrium). This paper argues that the means to establish this balance has been introduced by CSP. Extensions to the CSP approach could allow this balance to be adjusted to either side. For example the user may opt into more intrusive guidance that in the most extreme could result in a fully personalised website. This is possible due to the approach integrating directly with the website and therefore enabling deeper personalization scenarios. The main challenge related to increased intrusiveness is control. This paper argues that the control of intrusiveness should be based on the user's information needs. This however may conflict with the website publisher's economic need. An example of this conflict could be a user seeking full intrusiveness i.e. a fully guided and personal website, but not seeking any links that are of commercial benefit to the website publisher (such as promoted links, advertisement etc.). To overcome this shortcoming an adjustable or negotiable element could be introduced to the service that indicates the level of intrusiveness a user seeks and the level of invasiveness a website publisher allows.

Mobile and localization

The service-based approach ensures that CSP can be applied on websites rendered on a mobile device. This assists mobile use cases in which a user browses CSP enabled websites on a stationary desktop and then continues to browse the CSP enabled websites on a mobile device. This seamless Cross-Device Personalization can be extended by integrating CSP deeper into the mobile device. This would allow the CSP service to receive information about the user's location, and potentially allow the integration with other app information to enrich the user's cross-site profile and cross-site personalization. This extension might also include using the user's cross-site profile to provide location aware recommendations or by providing different apps with an informed decision.

A further, potentially mobile related, future work is in the context of localization. Specifically the application of CSP across different language sets. This would allow users to avail of CSP assistance on websites hosting content in different languages. An example of this would be in the tourism domain. Here a Non-English speaking user may seek to browse websites hosting in a local language and related to a specific hobby. The user could use CSP to seek visual assistance on English content-based websites, e.g. a tourist location, that match the user's hobby. To enable Cross-Lingual CSP two approaches could be studied. The first would translate all source (e.g. French) language based websites into one target language (e.g. English) and extract terms only from the target language corpus. The other, opposite approach, would be to extract terms in the source languages (e.g. French) and then translate the terms into one target language (e.g. English). Obviously the target language in both cases would relate to the

user's mother tongue and/or language preference. Future work in this space relies strongly on the investigation of semi-automated machine translation and multi-language term identification.

Web Narratives

One of the main concepts introduced by CSP is the notion of an *informed decision* that allows websites to provide navigational assistance without compromising the user's privacy.

The main focus of the approach introduced in this paper is focused on assisting the user in exploration and fact gathering based browsing across independently hosted websites. However, in tasks that require a more structured guidance and which may underlie a specific goal or objective, such as informal learning, the approach can be extended towards a more narrative driven decision process. The notion of introducing a narrative has been successfully applied in Adaptive Hypermedia Systems (AHS) traditionally focused on the area of e-Learning [52]. Within AHS the usage of a narrative allows the sequencing of course material based on the course description of a pedagogical expert. This narrative can then be used to populate a course with content based on pre-defined concepts. By overlaying the content concepts with user model concepts a balance between user preferences (e.g. learning goals, current knowledge etc.) and narrative preferences (a defined sequence of the course) [65].

7. ACKNOWLEDGMENTS

This research is supported by the Science Foundation Ireland (Grant 07/CE/I1142) as part of the Centre for Next Generation Localization (www.cngl.ie) at Trinity College Dublin. We would like to acknowledge the contributions made by Symantec and Fáilte Ireland in relation to providing content.

8. REFERENCES

- [1] D. O. Case, *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*. Emerald Group Publishing, 2012.
- [2] H. F. Korth and A. Silberschatz, "Database research faces the information explosion," *Commun ACM*, vol. 40, no. 2, pp. 139–142, Feb. 1997.
- [3] M. D. Kirchhoff, "Enaction: Toward a New Paradigm for Cognitive Science," *Philos. Psychol.*, vol. 26, no. 1, pp. 163–167, 2013.
- [4] Y.-C. Chen, R. A. Shang, and C. Y. Kao, "The effects of information overload on consumers' subjective state towards buying decision in the internet shopping environment," *Electron. Commer. Res. Appl.*, vol. 8, no. 1, pp. 48–58, Jan. 2009.
- [5] P. Epstein, "Beyond information: Exploring patients' preferences," *JAMA J. Am. Med. Assoc.*, vol. 302, no. 2, pp. 195–197, Jul. 2009.
- [6] A. Edmunds and A. Morris, "The problem of information overload in business organisations: a review of the literature," *Int. J. Inf. Manag.*, vol. 20, no. 1, pp. 17–28, Feb. 2000.
- [7] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Comput. Networks ISDN Syst.*, vol. 30, no. 1–7, pp. 107–117, Apr. 1998.
- [8] A. Broder, "A taxonomy of web search," *SIGIR Forum*, vol. 36, no. 2, pp. 3–10, Sep. 2002.
- [9] M. G. Noll and C. Meinel, "Web Search Personalization Via Social Bookmarking and Tagging," in *The Semantic Web*, vol. 4825, K. Aberer, K.-S. Choi, N. Noy, D. Allemang, K.-I. Lee, L. Nixon, J. Golbeck, P. Mika, D. Maynard, R. Mizoguchi, G. Schreiber, and P. Cudré-Mauroux, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 367–380.
- [10] P. Heymann, G. Koutrika, and H. Garcia-Molina, "Can social bookmarking improve web search?," in *Proceedings of the international conference on Web search and web data mining*, New York, NY, USA, 2008, pp. 195–206.
- [11] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 734–749, Jun. 2005.
- [12] S. S. Anand and B. Mobasher, "Intelligent Techniques for Web Personalization," in *Intelligent Techniques for Web Personalization*, vol. 3169, B. Mobasher and S. S. Anand, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 1–36.
- [13] M. Gao, K. Liu, and Z. Wu, "Personalisation in web computing and informatics: Theories, techniques, applications, and future research," *Inf. Syst. Front.*, vol. 12, no. 5, pp. 607–629, Jul. 2009.
- [14] M. Speretta and S. Gauch, "Personalized search based on user search histories," in *The 2005 IEEE/WIC/ACM International Conference on Web Intelligence*, 2005. Proceedings, 2005, pp. 622–628.
- [15] A. Dieberger, P. Dourish, K. Höök, P. Resnick, and A. Wexelblat, "Social navigation: techniques for building more usable systems," *interactions*, vol. 7, no. 6, pp. 36–45, Nov. 2000.
- [16] P. Brusilovsky and M. T. Maybury, "From adaptive hypermedia to the adaptive web," *Commun ACM*, vol. 45, no. 5, pp. 30–33, 2002.
- [17] F. Abel, E. Herder, G.-J. Houben, N. Henze, and D. Krause, "Cross-system user modeling and personalization on the social web," *User Model. User-Adapt. Interact.*, vol. 23, no. 2–3, pp. 169–209, 2013.
- [18] H. A. Feild, J. Allan, and R. Jones, "Predicting searcher frustration," in *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, Geneva, Switzerland, 2010, pp. 34–41.
- [19] E. Pariser, *The filter bubble: What the Internet is hiding from you*. Penguin Press HC, 2011.
- [20] T. O'Callaghan, "Eli Pariser: The dark side of web personalisation," *New Sci.*, vol. 211, no. 2822, p. 23, Jul. 2011.
- [21] B. Mobasher, R. Cooley, and J. Srivastava, "Automatic personalization based on Web usage mining," *Commun ACM*, vol. 43, no. 8, pp. 142–151, Aug. 2000.
- [22] K. Keenoy and M. Levene, "Personalisation of Web Search," in *Intelligent Techniques for Web Personalization*, vol. 3169, B. Mobasher and S. Anand, Eds. Springer Berlin / Heidelberg, 2005, pp. 201–228.
- [23] M. D. Mulvenna, S. S. Anand, and A. G. Büchner, "Personalization on the Net using Web mining: introduction," *Commun ACM*, vol. 43, no. 8, pp. 122–125, Aug. 2000.
- [24] S. Liewehr and M. Laplante, "Understanding Best Practices for Profiling, Personalizing, and Targeting Next Generation Engagement," *Gilbane*, Aug. 2011.
- [25] N. Nanas, A. Roeck, and M. Vavalis, "What Happened to Content-Based Information Filtering?," in *Advances in Information Retrieval Theory*, vol. 5766, L. Azzopardi, G. Kazai, S. Robertson, S. Rüger, M. Shokouhi, D. Song, and

- E. Yilmaz, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 249–256.
- [26] N. J. Belkin and W. B. Croft, "Information filtering and information retrieval: two sides of the same coin?," *Commun ACM*, vol. 35, no. 12, pp. 29–38, Dec. 1992.
- [27] M. J. Rudd and J. Rudd, "The Impact of the Information Explosion on Library Users: Overload or Opportunity?," *J. Acad. Librariansh.*, vol. 12, no. 5, pp. 304–06, 1986.
- [28] L. Sweeney, Zayatz, P., and Doyle, J., *Information Explosion. Confidentiality, Disclosure, and Data Access: Theory and Practical Applications for Statistical Agencies*. Urban Institute, 2001.
- [29] J. Sweller, "Cognitive load during problem solving: Effects on learning," *Cogn. Sci.*, vol. 12, no. 2, pp. 257–285, 1988.
- [30] A. Edmunds and A. Morris, "The problem of information overload in business organisations: a review of the literature," *Int. J. Inf. Manag.*, vol. 20, no. 1, pp. 17–28, Feb. 2000.
- [31] H. Berghel, "Cyberspace 2000: dealing with information overload," *Commun ACM*, vol. 40, no. 2, pp. 19–24, Feb. 1997.
- [32] D. M. Edwards and L. Hardman, "Lost in hyperspace: cognitive mapping and navigation in a hypertext environment," pp. 90–105, 1999.
- [33] P. Brusilovsky, "Adaptive Hypermedia," *User Model. User-Adapt. Interact.*, vol. 11, no. 1–2, pp. 87–110, 2001.
- [34] C. Bailey, W. Hall, D. E. Millard, and M. J. Weal, "Towards Open Adaptive Hypermedia," in *AH*, 2002, pp. 36–46.
- [35] Conlan and Wade, "Evaluation of APeLS—an adaptive eLearning service based on the multi-model, metadata-driven approach," in *Adaptive Hypermedia and Adaptive Web-Based Systems*, 2004, pp. 291–295.
- [36] P. Brusilovsky and N. Henze, "Open Corpus Adaptive Educational Hypermedia," *Adapt. Web*, vol. 4321/2007, pp. 671–696, 2007.
- [37] F. Bohnert and I. Zukerman, "Non-intrusive personalisation of the museum experience," *User Model. Adapt. Pers.*, pp. 197–209, 2009.
- [38] K. Koidl, O. Conlan, and V. Wade, "Non-Invasive Adaptation Service for Web-based Content Management Systems," in *International Workshop on Dynamic and Adaptive Hypertext: Generic Frameworks, Approaches and Techniques*, Torino, Italy, 2009.
- [39] B. Berendt and M. Spiliopoulou, "Analysis of navigation behaviour in web sites integrating multiple information systems," *VLDB J.*, vol. 9, no. 1, p. 56, 2000.
- [40] J. Kay, *Scrutable Adaptation: Because We Can and Must*. 2006.
- [41] S. Y. X. Komiak and I. Benbasat, "The effects of personalization and familiarity on trust and adoption of recommendation agents," *Mis Q.*, pp. 941–960, 2006.
- [42] J. Park and J. Kim, "Effects of contextual navigation aids on browsing diverse Web systems," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, The Hague, The Netherlands, 2000, pp. 257–264.
- [43] J. Brooke, "SUS-A quick and dirty usability scale," *Usability Eval. Ind.*, vol. 189, p. 194, 1996.
- [44] D. Hauger, A. Paramythis, and S. Weibelzahl, "Using browser interaction data to determine page reading behavior," *User Model. Adapt. Pers.*, pp. 147–158, 2011.
- [45] A. Paramythis, S. Weibelzahl, and J. Masthoff, "Layered evaluation of interactive adaptive systems: framework and formative methods," *User Model. User-Adapt. Interact.*, vol. 20, no. 5, pp. 383–453, Dec. 2010.
- [46] M. S. Bernstein, M. S. Ackerman, E. H. Chi, and R. C. Miller, "The trouble with social computing systems research," in *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, 2011, pp. 389–398.
- [47] R. Kohavi, A. Deng, B. Frasca, R. Longbotham, T. Walker, and Y. Xu, "Trustworthy online controlled experiments: five puzzling outcomes explained," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2012, pp. 786–794.
- [48] J. Lehmann, M. Lalmas, R. Baeza-Yates, and E. Yom-Tov, "Networked User Engagement."
- [49] F. Abel, N. Henze, E. Herder, and D. Krause, "Interweaving Public User Profiles on the Web," in *UMAP*, 2010, pp. 16–27.
- [50] Linden, Greg, Brent Smith, and Jeremy York. "Amazon.com recommendations: Item-to-item collaborative filtering." *Internet Computing*, IEEE 7.1 (2003): 76-80.
- [51] M. Feuz, M. Fuller, and F. Stalder. *Personal Web Searching in the Age of Semantic Capitalism: Diagnosing the Mechanisms of Personalisation*. First Monday, 16(2), 2011.
- [52] Conlan, Owen, and Vincent P. Wade. "Evaluation of APeLS—an adaptive eLearning service based on the multi-model, metadata-driven approach." *Adaptive Hypermedia and Adaptive Web-Based Systems*. Springer Berlin Heidelberg, 2004.
- [53] Michael J. Pazzani and Daniel Billsus. 2007. Content-based recommendation systems. In *The adaptive web*, Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl (Eds.). Lecture Notes In Computer Science, Vol. 4321. Springer-Verlag, Berlin, Heidelberg 325-341.
- [54] Sieg, Ahu, Bamshad Mobasher, and Robin Burke. "Web search personalization with ontological user profiles." *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management*. ACM, 2007.
- [55] Ginige, Athula, and San Murugesan. "Web engineering: An introduction." *MultiMedia*, IEEE 8.1 (2001): 14-18.
- [56] Rossi, Gustavo, and Daniel Schwabe. "Modeling and implementing web applications with OOHDM." *Web engineering: Modelling and implementing Web applications*. Springer London, 2008. 109-155.
- [57] Koch, Nora, et al. "UML-based web engineering." *Web Engineering: Modelling and Implementing Web Applications*. Springer London, 2008. 157-191.
- [58] Pastor, Oscar, et al. "Conceptual modelling of web applications: the OOWS approach." *Web Engineering*. Springer Berlin Heidelberg, 2006. 277-302.
- [59] Ceri, Stefano, Piero Fraternali, and Aldo Bongio. "Web Modeling Language (WebML): a modeling language for designing Web sites." *Computer Networks* 33.1 (2000): 137-157.
- [60] Acerbis, Roberto, et al. "Webratio 5: An eclipse-based case tool for engineering web applications." *Web Engineering*. Springer Berlin Heidelberg, 2007. 501-505.
- [61] Avison, David, and Guy Fitzgerald. *Information systems development: methodologies, techniques and tools*. McGraw Hill, 2003.
- [62] Isakowitz, Tomas, Michael Bieber, and Fabio Vitali. "Web information systems." *Communications of the ACM* 41.7 (1998): 78-80.

- [63] Komiak, Sherrie YX, and Izak Benbasat. "The effects of personalization and familiarity on trust and adoption of recommendation agents." *Mis Quarterly*(2006): 941-960.
- [64] Wood, Jason, Helen Wright, and Ken Brodie. "Collaborative visualization." *Visualization'97, Proceedings. IEEE*, 1997.
- [65] Abbing, Jana, and Kevin Koidl. "Template approach for adaptive learning strategies." (2006).