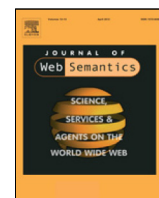




Contents lists available at ScienceDirect

Web Semantics: Science, Services and Agents on the World Wide Web

journal homepage: www.elsevier.com/locate/websem

Evaluating the impact of semantic technologies on bibliographic systems: A user-centred and comparative approach



Mariano Rico ^{a,*}, Daniel Vila-Suero ^a, Iuliana Botezan ^b, Asunción Gómez-Pérez ^a

^a *Ontology Engineering Group, Universidad Politécnica de Madrid, Spain*

^b *Librarianship and Information Science Department, Faculty of Information Science, Universidad Complutense de Madrid, Spain*

ARTICLE INFO

Article history:

Received 2 October 2017

Received in revised form 6 March 2019

Accepted 11 March 2019

Available online 19 March 2019

Keywords:

User-centred evaluation

Usability

Cultural heritage

Digital humanities

Linked data

Ontologies

ABSTRACT

Semantic and linked-data technologies are currently used by several cultural heritage institutions to make their content available through the Web. Although these technologies are heavily oriented towards data reuse and integration, one clear benefit highlighted by recent literature is the enhancement of human cultural consumption and user experience through the development of novel cultural end-user applications like Online Public Access Catalogues (OPACs). However, to the best of our knowledge, studies into the impact of these technologies on end-user applications are scarce. In order to address this lack, we report the results of two within-group user-centred studies of two online bibliographic systems in a realistic setting – using a widely deployed OPAC and its counterpart linked-data based system, *datos.bne.es*. The results of our first within-group study show that users of the system based on linked data required significantly less time and visited fewer pages to complete a typical search and retrieval activity. Additionally, the results of our user satisfaction tests also provided significantly better results for this new system. These results are consistent with the hypothesis that semantic technologies applied to library catalogues provide an enhancement that helps satisfy users' information needs.

© 2019 Elsevier B.V. All rights reserved.

1. Motivation

Cultural heritage organisations are starting to employ semantic and linked-data methods and technologies [1] to facilitate access to cultural and humanities content through the Web: European [2], the Library of Congress of the United States of America, the Digital Public Library of America, Worldcat, and *data.bnf.fr* [3] are prominent examples of this trend. Moreover, a number of ontologies are being designed and developed by standardisation bodies such as the IFLA (International Federation of Library Associations), the Joint Steering Committee for Development of RDA (Resource Description and Access), and the Bibliographic Framework initiative (BIBFRAME) to support cultural data modelling on the Web.

These methods, technologies and models come with the promise of enhanced access, user experiences and reuse of cultural heritage assets on the Web. However, to the best of our knowledge, little research has been undertaken into the impact of these novel approaches, especially regarding real-world uses

and tasks such as those related to the digital humanities and information sciences.

The goal of this paper is to study and evaluate a variety of aspects related to information retrieval, navigation of cultural content, and user experience, in a controlled but realistic scenario. We compared two production systems: the Online Public Access Catalogue (OPAC), and the linked-data based service (*datos.bne.es*) offered by the National Library of Spain (Biblioteca Nacional de España, BNE). The former is the main entry point to the collections held by the BNE and has a user base of more than 500,000 users per month, the latter – described in more detail in Section 2 – was released in November 2014 and has shown a steady growth in user numbers since its release (with more than 50,000 users per month since January 2015). These systems are of special relevance due to the social impact of the data holder, and the size, diversity and cultural significance of the collections the library provides access to. Furthermore, these two systems represent a suitable object of study because

- they enable access to the same underlying catalogue composed of more than 4 million MARC 21 bibliographic records, more than 3 million MARC 21 authority records (e.g., providing personal information, classification information, etc.), and to more than 100,000 digitised materials. In other words, both systems manage, process and make available the same amount and type of information.

* Correspondence to: Universidad Politécnica de Madrid, Campus de Montegancedo, sn, Boadilla del Monte, 28660 Madrid, Spain.

E-mail addresses: mariano.rico@fi.upm.es (M. Rico), dvila@fi.upm.es (D. Vila-Suero), ibotezan@pdi.ucm.es (I. Botezan), asun@fi.upm.es (A. Gómez-Pérez).

- they provide comparable functionalities in terms of information access, such as a search engine across different types of entities (e.g. authors, works, translated works, or subjects), navigation between different types of connected entities (such as the author of a work and their work). Yet, they take a completely different approach to information navigation, visualisation and retrieval.
- the OPAC system is a standard, widely deployed, and established mechanism to provide online access to library catalogues and represents current practice within the library domain.¹ The *datos.bne.es* system, as we describe in Section 2, applies semantic and linked-data technologies extensively, and is comparable both in scale and applied semantic techniques to other pertinent systems such as *data.bnf.fr* or the German National Library's linked-data services.²

In order to evaluate these systems we carried out a usability test. The ISO/IEC 9126-4 standard [4] defines usability as “the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”. Therefore, from this we can infer that a usability test should include three measurements: efficiency, effectiveness and user satisfaction. In this study, the first two were measured using a task-solving experiment that focused on content navigation and retrieval (see Section 4). User satisfaction was measured via a user-satisfaction survey (see Section 5).

Section 2 describes the main features of the *datos.bne.es* system. Section 3 describes the experiments' methodology. Section 4, the test environment. Finally, we discuss the related work in Section 6 and draw a conclusion in Section 7.

2. *datos.bne.es*

The *datos.bne.es* system is the result of a project that began in 2011 with the goal of investigating the application of semantic and linked-data technologies to the National Library of Spain's catalogue. Since 2011, the project has passed three milestones.

The first milestone was the transformation of a subset of the catalogue into RDF (Resource Description Framework), and modelled with the ontologies developed by the IFLA [5]. During this phase, we explored different techniques for both mapping and extracting relationships – and entities – from MARC 21 format records.

The second milestone was the generation of relationships and entities from a significant part of the catalogue following linked-data principles³ – and their subsequent publication. The main result of this phase was the release of a large and highly interlinked dataset under a CC0 1.0 Universal Public Domain Dedication license [6]. The dataset was queryable using either the SPARQL language through a public endpoint,⁴ or through a standard linked data front-end, Pubby,⁵ which provided access to the data in different formats using content-negotiation. During this phase, we improved and extended the methodology and mapping and extraction techniques, and built a tool, *Marimba*, that utilises the knowledge of cataloguing experts during the mapping process [7].

The last milestone of the project was the release of the service analysed in this paper. This version of the service introduced the following major improvements:

¹ The OPAC system used by BNE is powered by the Symphony commercial software used by more than 20,000 libraries (<http://www.sirsidynix.com/products/symphony>).

² Available at http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkeddata_node.html.

³ <http://www.w3.org/DesignIssues/LinkedData.html>.

⁴ <http://datos.bne.es/sparql>.

⁵ The front-end used by projects like the Spanish DBpedia.

- The complete catalogue was transformed and interlinked. It consequently covered 4,784,303 bibliographic records, 3,083,671 authority records, and generated 143,153,218 unique RDF triples. Moreover, the number of *owl:sameAs* links to external datasets was significantly increased – to 1,395,108 links. Additionally, 108,834 links to digital materials were added.
- The data is modelled using an integrated ontology, based on more than ten different bibliographic ontologies. This rich ontology has been published following the linked data principles and is publicly available.⁶ We describe the ontology in more detail in Section 2.1
- An end-user interface was developed to give access to the vast numbers of interconnected entities. This user interface was built exclusively using linked data and leveraged the data connectivity and the underlying ontology to index, present, and arrange information – as we describe in Section 2.2.

2.1. BNE ontology

A core component of the *datos.bne.es* system is its underlying ontology. The BNE ontology has been built collaboratively by cataloguing experts and ontology engineers using the *Marimba* tool. The tool articulates a mapping process where experts can define mapping rules using different domain ontologies; these are later incorporated into an integrating ontology and then used to create instances in accordance with linked-data best practice. In particular, the BNE ontology builds on the IFLA FRBR model (Functional Requirements for Bibliographic Records), although it reuses several terms from other ontologies such as the RDA ontologies, Dublin Core, and the Bibo ontology.

The FRBR model, developed by the library community, proposes an entity-relation conceptual model to improve the alignment between bibliographic concepts and the expected use of library information. Since its creation, there have been several efforts to model bibliographic data and to study the impact on bibliographic systems – as described in Section 6. The *datos.bne.es* project was one of the first projects to comprehensively apply the FRBR ontology to a large catalogue [5], and since then we have been widening the ontology. The BNE ontology reuses the core classes of FRBR (i.e., Person, Corporate Body, Work, Expression, Manifestation, and Item) but incorporates certain additional properties and relationships⁷ into the model, in order to broaden the coverage of such a large library catalogue. Fig. 1 shows the core classes and object properties of the BNE ontology; the reader can see how we relate and interconnect core FRBR concepts. Also, Table 1 presents the core terms of the ontology, their types (C indicates *class*, OP indicates *object property*) and their numbers of instances.

It might appear that the BNE systems have an unusually high proportion of authority records; however, it should be noted that, historically, BNE has dedicated resources to developing its authority file by including corporate bodies, persons, works and, to a lesser extent, expressions of these works. Most of the expressions in the catalogue correspond to translations of works. In a large number of cases, expressions corresponding with the original language of a work are not present in the authority catalogue. This is a known issue with FRBR-based systems, which the *datos.bne.es* system tackles by creating these expressions from existing information in authority and bibliographic records (e.g. in fields such as 008 and 245). This automatic enrichment process is described

⁶ <http://datos.bne.es/def/>.

⁷ For instance, a direct authorship relationship to connect persons and manifestations.

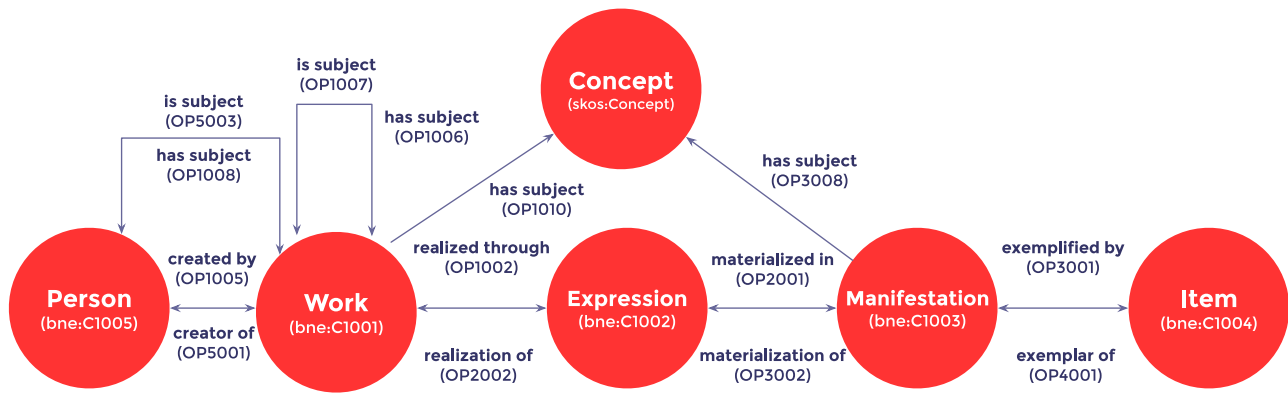


Fig. 1. Core of the BNE ontology. Each class and object property includes a descriptive label and its URI in the Compact URI syntax (see <http://www.w3.org/TR/curie/>), where the bne prefix corresponds to <http://datos.bne.es/def/>.

Table 1

Core terms of the BNE ontology, their types (C indicates *class*, OP indicates *object property*), their identifier (URI) expressed in the compact URI notation, and their number of instances.

| Element (Type) | URI | Instances |
|-----------------------|--------------|-----------|
| Work (C) | bne:C1001 | 1,958,242 |
| Expression (C) | bne:C1002 | 1,286,244 |
| Manifestation (C) | bne:C1003 | 4,784,303 |
| Item (C) | bne:C1004 | 108,834 |
| Person (C) | bne:C1005 | 1,238,723 |
| Corporate body (C) | bne:C1006 | 301,094 |
| Concept (C) | skos:Concept | 478,966 |
| creator of (OP) | bne:OP5001 | 1,085,055 |
| realised through (OP) | bne:OP1002 | 1,216,665 |
| materialised in (OP) | bne:OP2001 | 1,553,337 |
| exemplified by (OP) | bne:OP3001 | 108,834 |
| is subject (OP) | bne:OP3008 | 1,844,851 |
| is subject (OP) | bne:OP1007 | 10,143 |
| is subject (OP) | bne:OP1008 | 51,015 |
| is subject (OP) | bne:OP1010 | 13,253 |

in Vila et al. 2013 [7]. However, it is worth noting that even with this systematic mapping approach the current system is unable to connect every work to their corresponding expression. In the current data, this drawback is ameliorated by linking works directly to their corresponding bibliographic records. Finally, this is also partially remedied by the user-interface design, which presents works that only have expressions with the goal of presenting a homogeneous view of every work.

One of the objectives of this paper is to study the impact of this type of semantic modelling in a web context and as applied to large online catalogues; Section 4 explains the relation of the core components of the data model to the design of the tasks for the experiment — in order to evaluate the impact on end-users.

2.2. Ontology-driven user interface

The datos.bne.es user interface (UI) allows end-users to search and browse through information contained in the BNE catalogue. The BNE catalogue is structured according to established cataloguing rules such as ISBD (International Standard Bibliographic Description) and bibliographic standards such as MARC 21. Library catalogues are based on bibliographic records, which are to a great extent both structured and normalised, this has enabled the use of common “advanced search” options for years, such as filtering and field searching, topic browsing and searches using controlled vocabularies. Linked-data solutions and UIs can build on these features and provide the opportunity to explore new mechanisms for information navigation, retrieval and visualisation. In particular, the datos.bne.es linked-data UI makes extensive use of linked data’s graph structure and ontological model to

deliver enhanced access to information. We have classified it into these three types:

Search (UI-S). The datos.bne.es system follows an entity-centric approach to indexing, ranking and search. The search engine indexes properties describing entities that are modelled according to the BNE ontology (e.g. Persons, Works, Concepts). Ranking and relevance of results are based on weighted links across entities, and qualified by the object properties in the ontology. For authors, for instance, the more created works, the better ranked the author.

Navigation (UI-N). Information navigation is enabled exclusively by following qualified links from one instance to another (e.g. from an author to their work, then to the subject of a work). Each page describes an instance of the ontology, and is linked to other instances via its object properties.

Presentation (UI-P). The presentation of information is also derived directly from the ontology, in the sense that the main information blocks of each entity page and their order are determined by the semantic model (e.g. the author pages hold three major blocks defined by the most relevant object properties: creator of, contributes to, and is subject of).

Section 4, we explain how these features were explored in the user-centred experiment which investigated their impact on end-users.

3. Methodology

To evaluate effectiveness and efficiency, we compared the two systems in an end-user experiment with 72 participants. In this experiment, participants were presented with a ten-task scenario which they had to complete using one of the systems. Our experiment to evaluate satisfaction had fewer participants (36). The goal of these experiments was to evaluate a set of three initial hypotheses:

H1. A user will find and retrieve the required information significantly faster with the datos.bne.es system than with the traditional OPAC system.

H2. A user will find and retrieve the required information by visiting significantly fewer pages with the datos.bne.es system than with the OPAC system.

H3. User satisfaction as regards the user interface will be significantly better with datos.bne.es than OPAC.

These hypotheses share the underlying assumption that semantic technologies applied to library catalogues have a positive impact on users in terms of satisfying their information needs. To investigate these hypotheses, we designed a controlled user-centred experiment — described in the sections which follow.

3.1. Experimental setting

The experiments were conducted in a controlled environment with a selection of Library and Information Science students and Journalism undergraduates at the Universidad Complutense de Madrid. Each experiment divided randomly the participants into two groups. One group performed the experiment using the OPAC system (OPAC user group); the other, using *datos.bne.es* (*datos.bne.es* user group). We applied an experimental blind where the participants did not know which system they would be using in advance.

Although the experiments were conducted over several sessions, all participants worked within the same technical setting and in the same computer room and where all the computers had similar technical specifications. Each participant worked within the same scenario – a contextualised data-retrieval activity – and performed exactly the same tasks, but the *OPAC user group* used the system at <http://catalogo.bne.es> and the *datos.bne.es user group* used the system at <http://datos.bne.es>.

We measured two quantitative variables: total time and number of pages to complete each task; and one subjective variable: satisfaction (on a five-point Likert scale). These measurements were the dependent variables for the independent variables of our study: the two systems being compared.

The experiment was conducted using an online application specifically developed for the experiment, as shown in Fig. 2.

3.2. Participants

We selected students from the Journalism and Information and Documentation degree schemes as they were already familiar with the traditional BNE OPAC interface, and were also used to information-seeking tasks involving bibliographic searches in different libraries and library catalogues. Information and Documentation students were not only familiar with traditional bibliographic tools as end-users, but also as practitioners. They were knowledgeable about the records' structure, had skills in the use of controlled vocabularies and even possessed bibliographic record management skills.

Although the evidence we have gathered is not conclusive, we do think that having put *datos.bne.es* at a certain disadvantage strengthens our conclusions about users' preference for *datos.bne.es* over OPAC. Users knew how to use with the OPAC beforehand and could take advantage of the advanced search features; at the same time, they had never used *datos.bne.es* before, which approaches the retrieval task quite differently.

The questionnaire collected two personal attributes (age and year of study) for each participant, and two attributes related to experience and skills with the systems being studied (OPAC and *datos.bne.es*). In order to establish experience levels,⁸ the participants indicated how many times they had used the systems previously on a 6-point scale: 0, [1–10], [10–20], [20–40], [40–60], or more than 60 times.

For the first experiment – aimed at evaluating effectiveness and efficiency – we had sufficient participants (72) to analyse variables such as age, year of study and previous experience. However, for the second experiment there were insufficient participants (36) to provide a detailed analysis of these aspects.

In the first experiment, the age of the 72 participants ranged from 18 to 58 years with an average age of 26.98 years (sd = 9.94). The age of participants within the OPAC ranged from 18 to 56 years and averaged 27.47 years (sd = 9.67 years). The age of

participants within *datos.bne.es* ranged from 18 to 58 years and averaged 26.5 years (sd = 10.32).

Focusing on year of study, overall the first experiment included undergraduate students in their first (19 participants), third (28), and fourth years (14), as well as postgraduate students (11). Divided by year of study, the OPAC included 10 participants in their first year, 12 in their third, 8 in their fourth and 6 were postgraduate students. The *datos.bne.es* had 9 participants in their first year, 16 in their third, 6 in their fourth and 5 were postgraduate students.

As regards previous experience, only 12.5% of participants had never used the OPAC system, approximately 15.3% had used it more than 60 times, 26.4% more than 40, and 52.8% less than 20. The *datos.bne.es* system was less well-known: approximately 28% of participants had never used it, and only 3.6% more than 20 times. In the OPAC, 19.4% of participants had used the OPAC system more than 60 times previously, 38.8% more than 20 times, and 47.2% between 1 and 20. Within the *datos.bne.es*, 88.9% of participants had used the system less than 10 times, and the remaining 11.1% between 10 and 40.

4. Task solving scenario: experiment and evaluation

In order to evaluate the quantitative variables mentioned in 3.1 (total time and number of pages required to complete each task), we designed a realistic scenario composed of ten consecutive, related tasks. The workflow of the experiment had three steps:

- Context-of-use questions. In order to understand the characteristics of each participant, the system asked them to fill in a form. Specifically, we focused on *personal attributes* (age and year of study), and *experience, knowledge and skills* (previous experience with the OPAC and *datos.bne.es* systems).
- Introduction to the scenario. The system introduced the scenario by providing a detailed description. This presented the scenario to the participants – a data-retrieval situation which could be realistically be asked of Journalism or Information and Documentation students.
- Tasks. The system presented the participant with a set of ten consecutive tasks in the same order. Each task was composed of a title and a description. For instance, the user was asked: *Find and retrieve the title of two works written by Santiago Ramón y Cajal*. The maximum time to complete each task was ten minutes. Once the participant had read the description, the experiment presented her with a countdown timer and the system screen to start the task. Once the user had finished the task she had to press a stop button. Following this, she was presented with a form to enter the answer to the question and a brief questionnaire for subjective assessment.

We used a between-subject design; that is to say, we divided the participants in two groups, in which each participant only used one tool, and all participants had to carry out the same tasks. We also considered a within-subject design, in which the participants test both systems; however, this would have had some drawbacks, like a carryover effect (practice on the first system conditions the second system) and fatigue [8]. Consequently, this was rejected in favour of the between-subject design. We checked that the two groups were similar (no skew) after randomly dividing the participants into two groups.

The scenario was based on the actual uses to which the OPAC system at the Complutense University was being put – information available as one of the authors is affiliated with the university. Thus, the scenario was not created with *datos.bne.es* in mind, but rather aimed to be a task that mimicked the kind

⁸ Unless otherwise is specified, consecutive ranges should be considered with a non inclusive upper limit. For instance, the ranges (10–20) and (20–40) should be considered as [10, 20) and [20, 40).

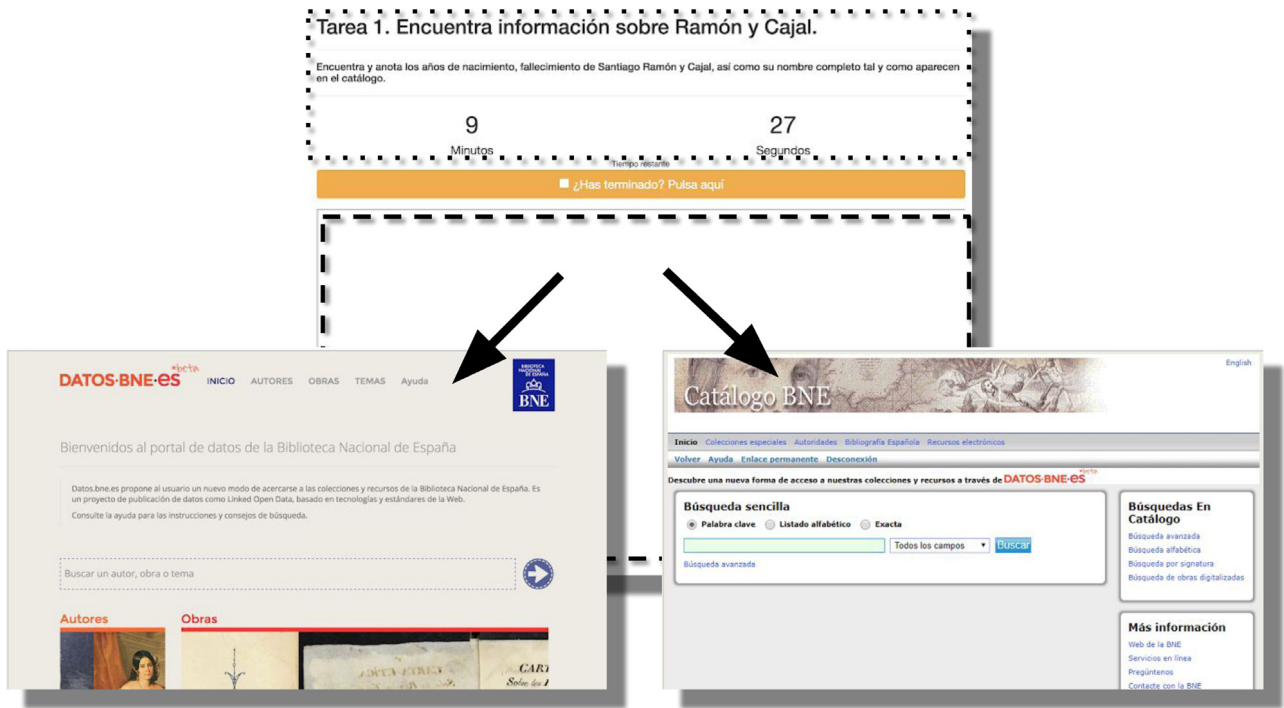


Fig. 2. Experimental setup for OPAC (right image) and datos.bne (left image). The experiments' framework is in the upper side image. The dashed box shows the frame in which the web application of each system was located. The dotted box shows the task id ("Tarea" in Spanish), the description of the task ("find information about Ramón y Cajal") and the time which remains to finish the task. Between the dotted box and the dashed box there is a wide button that participant had to press once the proposed task had finished.

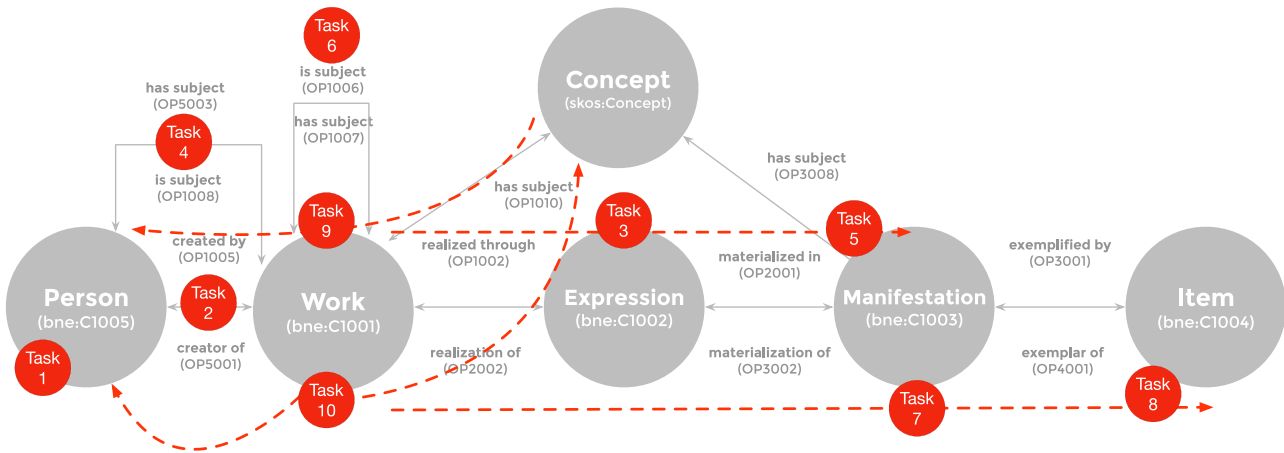


Fig. 3. Relationship between tasks and core terms of the ontology.

Table 2

Description of tasks comprising the scenario. A more detailed description can be found in [Appendix A](#).

| Task id | Description | Navigation | UI feature |
|---------|--|---|------------------|
| (1) | Find author information | {Person} | UI-S |
| (2) | Find works of an author | {Person, Work} | UI-P |
| (3) | Find work information | {Person, Work, Expression, Manifestation} | UI-N |
| (4) | Find a work about an author | {Work, Person} | UI-P |
| (5) | Find a translated version of a work | {Work, Expression, Manifestation} | UI-N |
| (6) | Find a work about another work | {Work, Work} | UI-S, UI-N |
| (7) | Find an edition of a translated work | {Work, Expression, Manifestation, Item} | UI-N, UI-P |
| (8) | Find a digital version of a work about an author | {Work, Expression, Manifestation, Item} | UI-S, UI-N, UI-P |
| (9) | Find information about a topic | {Concept, Work, Person} | UI-S, UI-N, UI-P |
| (10) | Find the topic and author of a work | {Work, Concept, Person} | UI-N |

of tasks students might be asked to carry out during their degree schemes. From the point of view of the user, the scenario presents an overall objective: *to collect information related to Santiago*

Ramón y Cajal, one of the most important Spanish scientists of the last century, for the documentation department of a media company that is preparing an article about the author.

This main objective was then divided into ten consecutive and descriptive tasks with different levels of complexity. The reason for dividing the activity into 10 predefined tasks was to be able to precisely measure efficiency and effectiveness for each of the different tasks, and therefore different aspects of the systems (e.g. searches involving translations of works or author information).

It is important to note that the BNE's OPAC system does provide mechanisms to browse information, allowing users to go from authors to bibliographic records and subjects. The fact that most of the users of the OPAC system managed to finish the complete set of tasks was probably due to this feature.

In the absence of a large enough space for all the participants to undergo the experiment at once, we divided the experiments into several sessions. Each session was held on a different day, and in each session there were participants from both OPAC and datos.bne. During each, we randomly divided the students into two groups and provided them with the URL of the system their group corresponded to.

Despite this randomisation, we nevertheless carried out a context-of-use analysis in order to assess the impact of participants' experience and skills on the results of the study (See Section 3.2).

Ingwersen and Järvelin [9] argue that there is a need for "a further development of the cognitive viewpoint for information seeking and retrieval by providing a contextual holistic perspective" in information retrieval research. User-centred approaches offer several advantages. For example, simulated work tasks can trigger participants' "simulated information needs" and work as "the platform against which situational relevance (usefulness) is assessed" [9]. Our experimental scenario provided participants with a context and realistic information need; the problems to be solved were specifically intended to make the situational relevance evident to the user.

With regard to the complexity of individual tasks, it is worth noting that the main aim of our user-centred experiment was the evaluation of a full and realistic exercise; therefore, there were several tasks that might be considered easily solved with either system (such as retrieving biographical information about an author), and other more complex tasks (e.g. finding the subject of a certain edition). Thus, our main hypotheses were about the full scenario of interrelated tasks, rather than the performance in individual tasks. Nevertheless, while the overall scenario was system-agnostic, to broaden the scope of the study we designed the tasks to evaluate key aspects that would presumably differentiate datos.bne.es from OPAC systems. Furthermore, the experiment aimed to keep a balance between a broad analytical scope, coverage of features, and maintaining a scenario which would appear coherent to the participants.

Each task was designed taking into account two different aspects: (1) the ontology terms, and (2) the UI features. Fig. 3 shows the relations between tasks and the ontology classes and properties presented in Section 2.1 graphically. Dashed red arrows indicate the expected path to complete the task, and the position of the red circle indicates the entity which the information was expected to be retrieved from. Additionally, the tasks accounted for the UI features described in Section 2.2. Table 2 summarises the 10 tasks, expected navigation, and related UI features. A complete description of the tasks is provided in Appendix A.

4.1. Results for time to complete all tasks

As regards our first hypothesis (H1), Table 3 shows the average time required by participants to complete all the tasks for each system. It also shows the average time required per task. Notice that, on average, users using datos.bne.es required 32% less time

Table 3

Average time (seconds) to complete all tasks, showing the average time for each task and the percentage differences relative to OPAC.

| Task ID | System | | ΔT relative to OPAC |
|----------|--------------|------|-----------------------------|
| | datos.bne.es | OPAC | |
| T_1 | 169 | 208 | -18.7% |
| T_2 | 184 | 234 | -21.4% |
| T_3 | 136 | 220 | -38.2% |
| T_4 | 205 | 261 | -21.5% |
| T_5 | 187 | 379 | -50.7% |
| T_6 | 279 | 349 | -20.1% |
| T_7 | 171 | 228 | -25.0% |
| T_8 | 238 | 272 | -12.5% |
| T_9 | 186 | 232 | -19.8% |
| T_{10} | 225 | 445 | -49.4% |
| Total | 1980 | 2828 | -30.0% |

to complete all the tasks than OPAC users. Some tasks, like T_5 and T_{10} , were performed in half the time (50%) when compared to OPAC.

A detailed breakdown of the total time each participant required is shown in Fig. 4(a). Each dot represents a participant. The "violin" figure shows the distribution of the responses, and the three error bars show the standard deviation (thin grey external bars), 95% confidence interval (thick grey middle bar), and standard error (thick black inner bar). We use these "lightsaber" bars⁹ to provide a graphical check of the statistical significance. Note that using this visualisation we can be very confident that the different values provided for these two systems are statistically significant; that is to say, not a mere statistical fluke. The statistical foundation of this visual approach can be found in Cumming et al. [10].

Non-parametric statistical methods such as ANOVA (with very restrictive requirements like normal distribution and homogeneous deviation), or t-student tests (less restrictive), provide a p-value that is easy to "cheat" when there is a small sample size. That is why we are interested in visualising the distribution form (violin plot) and the individuals (points) for both distributions. This graphical visualisation provides more information than the classical p-value measured with any statistical metric.

The right side of Fig. 4 (Fig. 4(b)) shows the number of pages each user required to complete all the tasks. On average, datos.bne.es users visited 40% fewer pages than OPAC users. The "lightsaber" error bars show a significant difference between the systems which is unlikely be due to a statistical fluke (Hypothesis H2).

We analysed the effect of user characteristics (e.g. age, year of study, experience). The most relevant findings were related to age and experience. Fig. 5 shows the time required to complete the proposed scenario broken down by age. We aggregated ages into three ranges. It is noticeable that, statistically, there was no significant difference for participants in the 17 to 22 and 22 to 30 year-ranges. However, people in the higher range (30 to 65) showed a significant difference: on average, participants in the 30-65 range needed 15% more time than younger participants (this applies to both datos.bne.es users and OPAC users).

For experience, Fig. 6 shows the time required to complete the tasks, aggregating participants by the number of times they said they had used the OPAC system. Although we had expected that the most experienced OPAC users would complete the tasks faster, we can see that there were no significant differences for any range of usage.

Fig. 7 shows a similar aggregation, but for experience with datos.bne.es. We had also expected better results for participants familiar with the new system, but the results did not provide statistical evidence for this.

⁹ See <http://datosbne.linkeddata.es/lightsabererrorbars.html>.

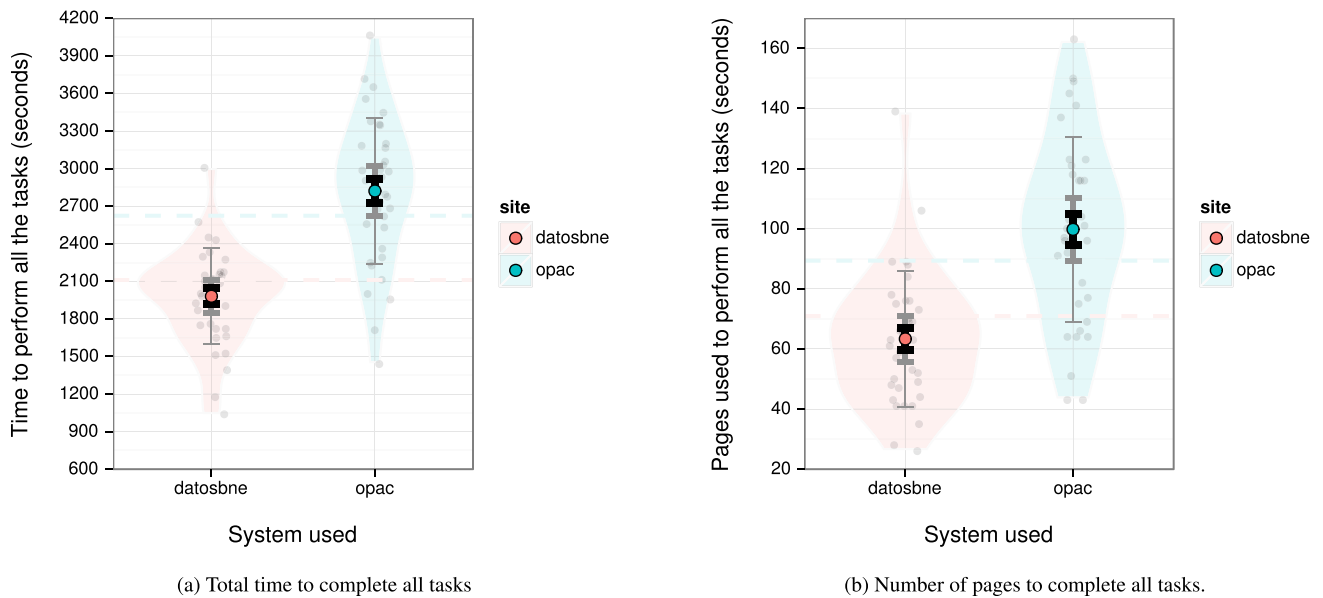


Fig. 4. Time and pages required to complete the tasks. Each dot is a user. As the 95% confidence intervals overlap less than 50%, the differences between systems are significant and we can reject that the possibility these are due to random fluctuations in the samples.

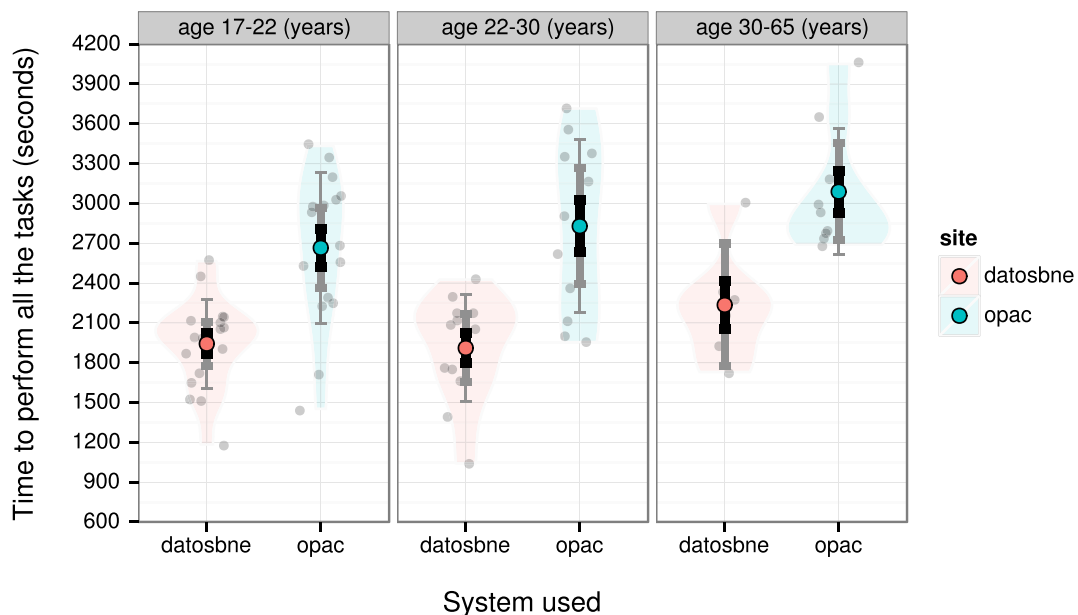


Fig. 5. Effect of the age. For the three ranges considered there is a statistically significant increment in the time required to perform all the tasks for OPAC. Also there is a statistically significant increment for both systems, datos.bne.es and OPAC, when comparing participants from the lower range (17–22) and the upper range (30–65).

4.2. Per-task results

This section discusses the results of each separate task for each given system, and in relation to our hypotheses about overall task completion. In particular, we were interested in studying how each task, and the participants' performance, were related to the semantic modelling (Section 2.1) and UI features (Section 2.2).

Overall, datos.bne.es performed better both in time and number of pages visited for every single task. Some differences were more prominent (Tasks 3, 5 and 10), while other – less complex – tasks like 1 and 2 showed improvement but not one which was significant. The task which was completed fastest on average with datos.bne.es was Task 3, with OPAC Task 1; while the slowest using datos.bne.es was Task 6, and with OPAC Task 10. For sample

dispersion by system, the results by task confirm that the OPAC presented a much more dispersed sample of results. However, there were some tasks in the datos.bne.es which presented a more dispersed distribution (especially Task 6).

As shown in Fig. 8(a), Tasks 5 and 10 were performed significantly faster using datos.bne.es. With respect to the UI, both tasks had a strong navigational component (UI-N) as they required the user to navigate across several entities. Results seem to show that with these kinds of complex navigational tasks users perform significantly better with datos.bne.es, due to the graph structure and underlying semantic modelling. Task 3 had similar characteristics (though to a lesser extent) and also had a strong navigational component. With regard to the impact of the semantic modelling, Tasks 3 and 5 shared common features, which were the use of the Work, Expression and Manifestation concepts, where a user starts

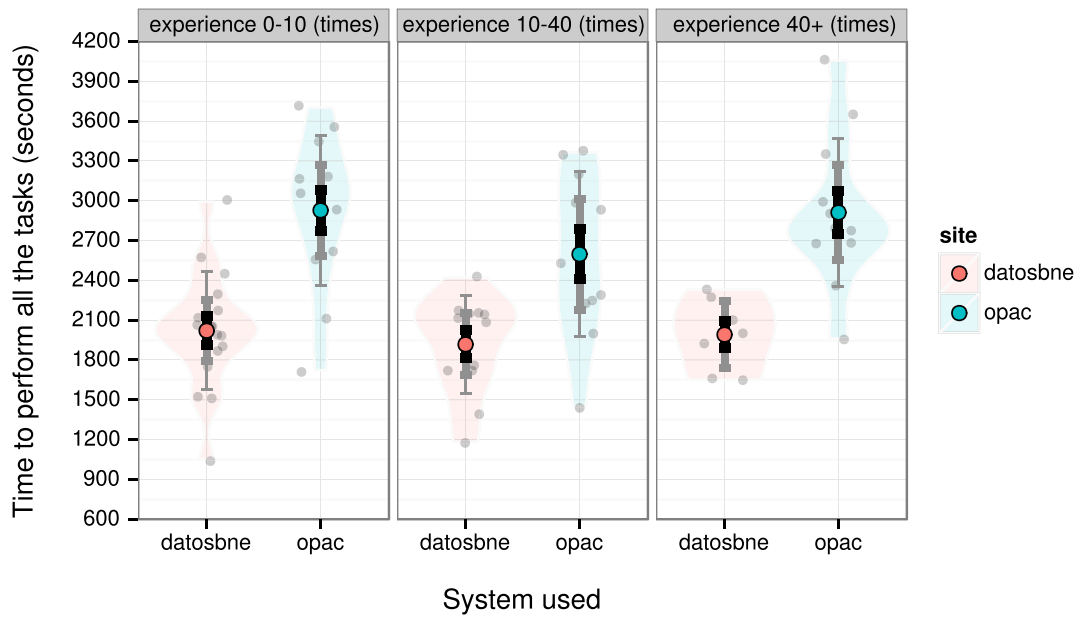


Fig. 6. Effect of experience on the OPAC system. There was no significant relation between speed and experience on the OPAC system.

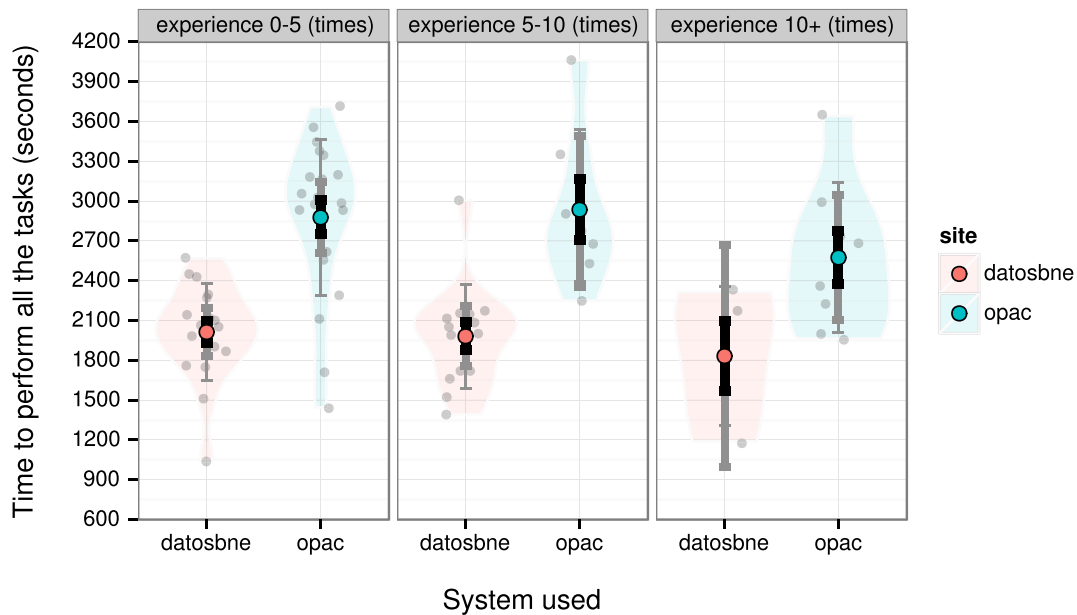


Fig. 7. Effect of experience on datos.bne. There was no significant relation between speed and experience on the datos.bne system.

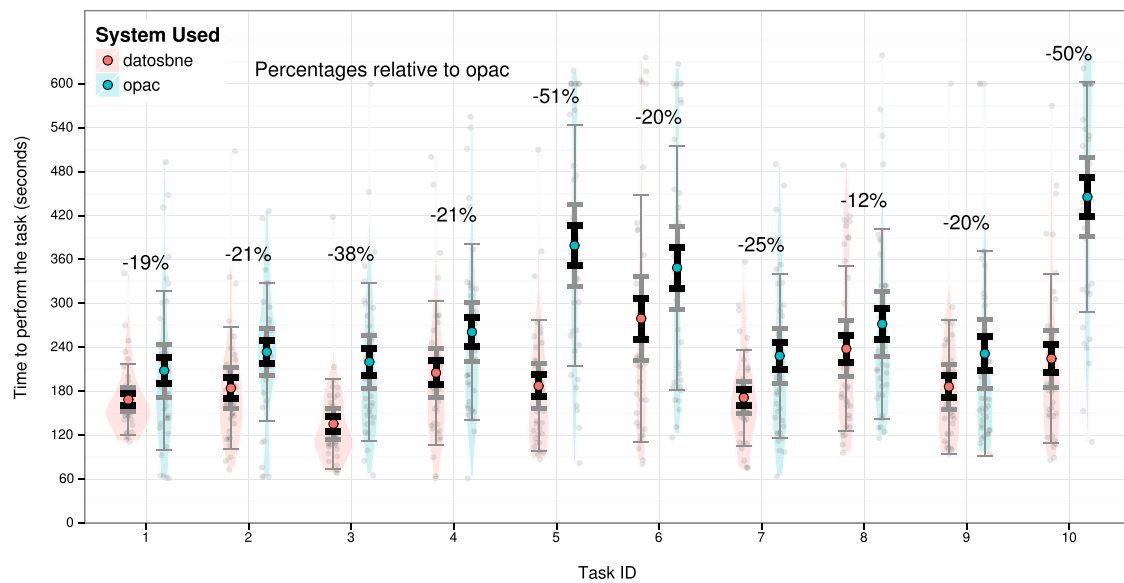
with a work and navigates to different versions and editions. However, the time differences between Tasks 3 and 5 seem to be a result of the language element (Task 5 requires the title of a work in English to be retrieved). It is worth noting that finding editions by languages could be reproduced in the OPAC system through the advanced search interface, but for the sample we studied this did not seem to aid users' task completion.

Finally, in Fig. 8(b), results show that there is not always a correlation between task completion time and number of pages visited. For instance, Task 1 and 2 have comparable distributions with regard to time but not with regard to pages.

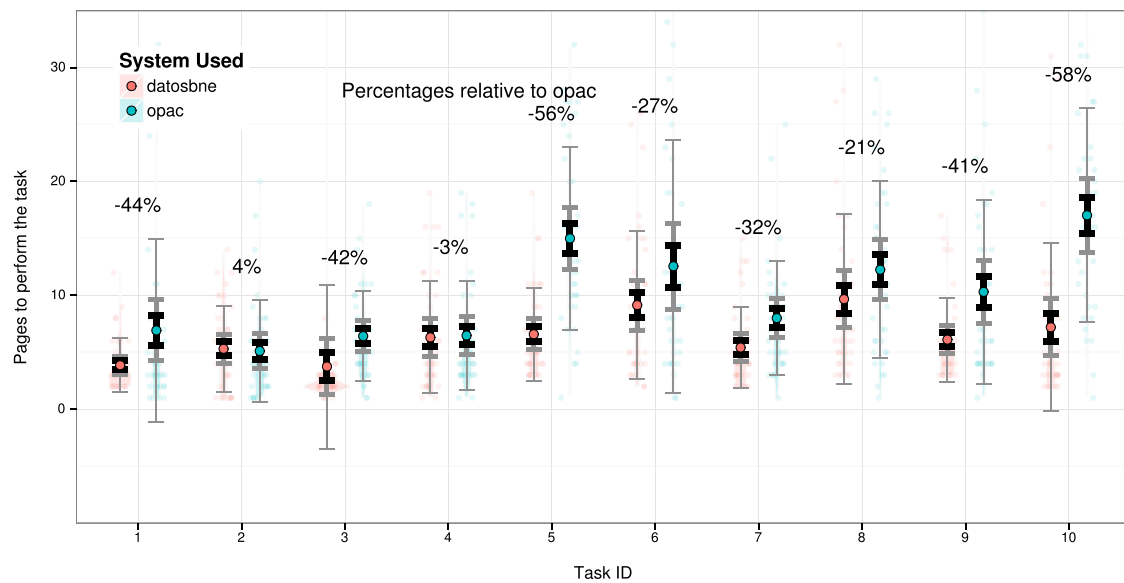
Regarding user characteristics, Fig. 9 is a detailed view of Fig. 8(a) showing each participant's year of study. We could not identify a significant difference for any year of study.

5. User-satisfaction evaluation

In order to measure user-interface satisfaction, two sets of 18 participants (independent of the participants described in Section 3.2) were selected from students from the Library and Information Science degree. One set evaluated the OPAC system, and the other datos.bne.es. The number of participants was a balance between precision and effort. In order to achieve a 90% confidence level for a given mean with error lower than 1%, at least 15 measurements are required [11]. This assumes intervals based on a normal population distribution for the mean, as we considered was the case here. Therefore, we recruited a slightly higher number of participants than this minimum.



(a) Differences in time required to complete all tasks



(b) Differences in number of pages visited to complete all tasks

Fig. 8. Differences in time required to achieve a task and pages visited for each system (OPAC and datos.bne). The error bars correspond to standard deviation (most external), 95% confidence interval, and standard error (most inner bars).

Participants were presented with the same tasks (Table 2) as the rest of participants but, once finished, a detailed questionnaire was presented to each participant (see Appendix B). Although there was no maximum time to finish the questionnaire, most users required around 20 min.

In order to evaluate user satisfaction, we used the standard “User Interface Satisfaction” test (QUIS) [12].

Valid responses to these questions are positive integers ranging from 0 (not satisfied at all) to 7 (completely satisfied). The results for OPAC and datos.bne.es are shown in Fig. 10. The average value for user satisfaction was 5.16 for OPAC and 5.48 for datos.bne.es, with a standard deviation of 0.35 and 0.49 respectively.

It is important to note that when QUIS was created in 1988 it was intended for expert user-interface designers and some questions could arguably be difficult for end-users to understand.

However, nowadays most of these questions are more understandable and, additionally, we made an effort to simplify the questions to aid end-user comprehension. We have used these simplifications in previous studies, like [13] and [14]. What we have improved in this paper is the presentation of data distributions. In our previous studies we only provided 2 error bars (standard error and standard deviation). Some reviewers requested more detailed information about the distributions because it is well known that anomalous distributions can produce low p-values. That is why we enhanced the graphics by adding the violin plot, the instances, and a method to establish visually whether two distributions are statistically significant. We consider this to be one of this paper’s valuable contributions; we make use of solid statistical methodology extracted from a reputable paper [10].

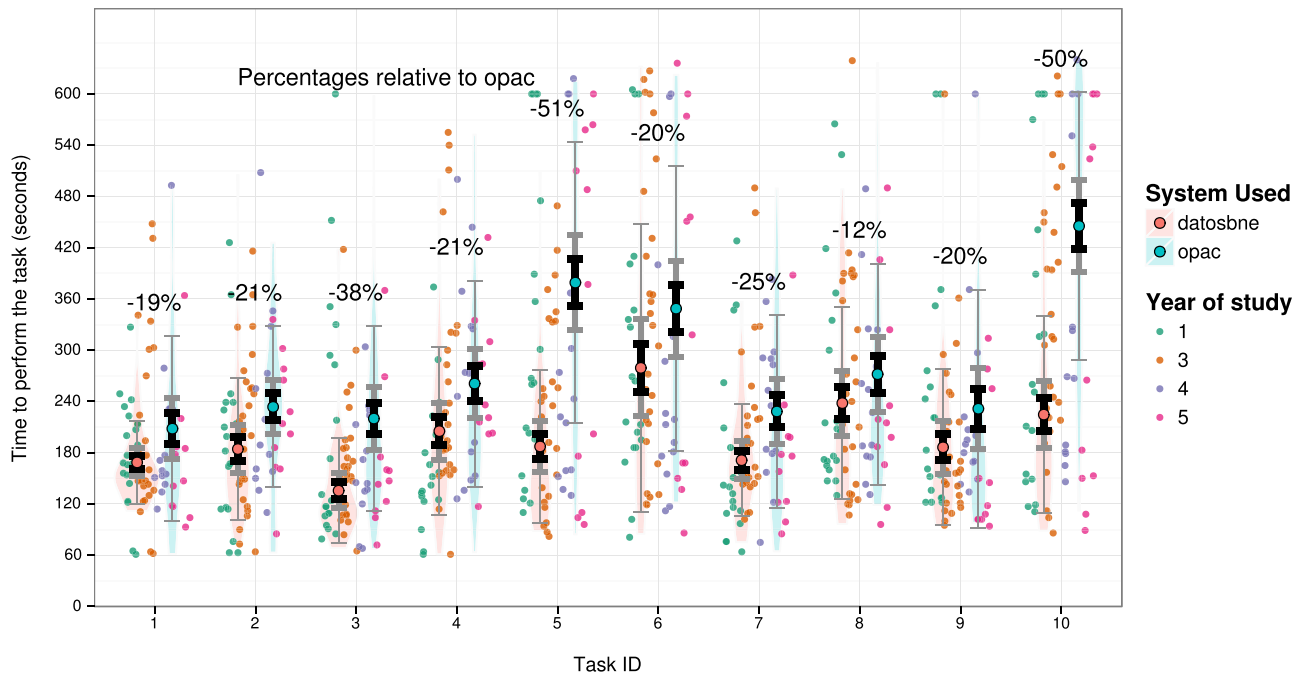


Fig. 9. Effect of year of study. We did not find any significant segmentation by year of study.

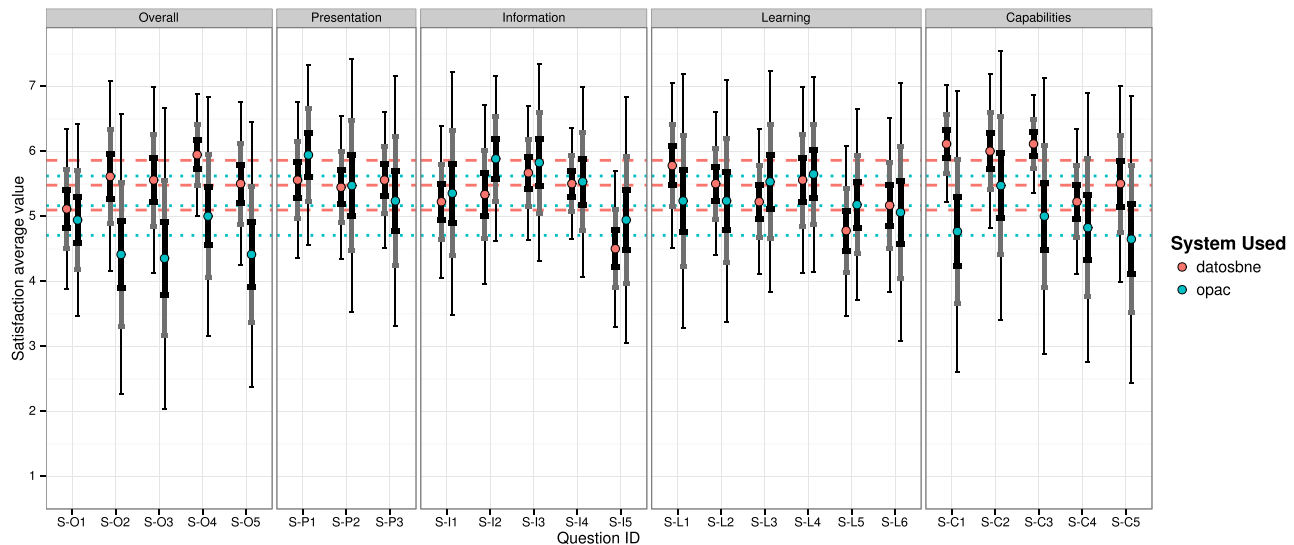


Fig. 10. Satisfaction test results. The horizontal lines show average and standard deviation for each system (computing all questions). Dashed lines are for datos.bne and dotted lines are for OPAC. On average, datos.bne presents better results than OPAC, but the differences cannot be considered significant when the bold error bars of the systems being compared overlap. The questions are available in [Appendix B](#).

5.1. Results

Fig. 10 shows that, on average, datos.bne.es has better user satisfaction values. If we focus on the five user-satisfaction question categories, we can see that datos.bne.es was rated better for 5/5 questions in the “Overall” category; there was no clear advantage for “Presentation”; datos.bne.es was scored lower compared to OPAC for “Information”; received better evaluations in 3/6 questions for “Learning”; and achieved better ratings in 5/5 questions for “Capabilities”. However, the lightsaber error bars show that in many cases there was not sufficient statistical evidence to confirm the better results obtained for datos.bne.es. We cannot reject the possibility these results were due to random error.

The free text questions permitted us to gather recommendations and suggestions from the participants. Most of these recommendations and suggestions will be implemented in the next version of the application. Positive aspects are in [Table 4](#) and negative aspects in [Table 5](#). We show aspects ranked by the percentage of participants who noticed that aspect. Only aspects with at least 10% are shown.

We also measured the strength of the comments (4 levels, from “Very positive” to “Not positive”), as shown in the lower part of each table. Users of datos.bne.es observed many more positive than negative aspects, and in general provided longer answers; aspects related to their interaction with the system were

Table 4

Most common **positive** aspects extracted from the satisfaction questionnaire's free text questions.

| OPAC | datos.bne.es | | |
|----------------------|--------------|-------------------------|-----|
| Advanced search | 39% | System easy to use | 30% |
| System easy to use | 22% | Results page display | 15% |
| Perceived speed | 17% | Intuitive search | 12% |
| Usefulness | 11% | Assistance in retrieval | 10% |
| Strength of comments | | | |
| Very positive | 54% | Very positive | 40% |
| Positive | 38% | Positive | 52% |
| Mildly positive | 0% | Mildly positive | 8% |
| Not positive | 8% | Not positive | 0% |

Table 5

Most common **negative** aspects extracted from the satisfaction questionnaire's free text questions.

| OPAC | datos.bne.es | | |
|-------------------------|--------------|-------------------------|-----|
| Relevance of results | 25% | System difficult to use | 18% |
| System difficult to use | 25% | Lack of advanced search | 15% |
| Assistance in retrieval | 12% | Web design flaws | 12% |
| Strength of comments | | | |
| Very negative | 31% | Very negative | 26% |
| Negative | 54% | Negative | 48% |
| Mildly negative | 8% | Mildly negative | 13% |
| Not negative | 8% | Not negative | 13% |

praised highly. OPAC users seem to have valued the advanced-search feature, which fitted with comments on the system's ease-of-use and the ability to complete tasks rapidly. Major negative issues reported by datos.bne.es users were related to their interaction with the system (difficult to use, web-design flaws, difficult-to-understand results page or difficulties accessing digital versions) or to the process of retrieval itself (lack of advanced search, confusing search box, problems with subject-matter retrieval).

Some of these issues are difficult to tackle, and we should conduct further tests to evaluate the effectiveness of the improvements that have already been made: an advanced search option was added after the experiment, subject matter retrieval was fixed and interaction with the search box improved. In general, datos.bne.es has improved significantly with respect to some of the most worrying issues – and still has room for further improvement. Conversely, notable problems reported by the OPAC group were not reported by datos.bne.es users. This seems to support our claim that users favour datos.bne.es over OPAC for most relevant aspects.

6. Related work

Borlund and Ingwersen [9,15] proposed an experiment based on a reduced number of users. Just as these authors did, we provided users with a plausible scenario to contextualise the information needs embedded within the tasks, and made the experiment a simulated work task. All of which served the purpose of getting the user to understand the goals of the research they were participating in. The Borlund and Ingwersen study compared two classical IR techniques, the “ranked output facility target” and the “exact match oriented Boolean Based quorum technique”, while our study compares two very different systems, a traditional OPAC and a novel tool, the BNE linked-data catalogue.

It appears that Borlund [16,17] was the first to test IS&R with real information needs, with an analysis of experimental results from a methodological point of view. Some of the methodological recommendations proposed by Borlund were later adopted by Ingwersen and Järvelin [9]. We have followed suit by setting

tasks that users could actually need to do in their everyday work; realistic tasks that motivate the test users to interact effectively with the system. They also highlight the need to permute the order of search jobs between the participants and that of pilot testing prior to actual testing. Ingwersen and Järvelin [9] suggest that the main advantage of this kind of study is the ability to measure results. Baeza-Yates and Ribeiro-Neto [18] add that this kind of experiment permits an “understanding [of the] dynamic characteristics of the user–system interaction that cannot be evaluated using static reference collections”. In addition, in our case we think that having been able to recruit actual users as test subjects to work on comparable systems, collections and near-identical tasks gives us a good base to generalise our results to the contexts and tasks that users utilise such systems in and for on an everyday basis.

Díaz et al. [19] compare user-centred and system-centred evaluation methods and conclude that user-centred evaluation yields “valuable data concerning the behaviour of the users with respect to issues such as document relevance or the relative importance attributed to different ways of personalization”. This validates the selection of user-centred evaluation methods such as those described in this paper.

Petrelli [20] analyses the role of user-centred information retrieval evaluation and describes it as “a flexible, adaptable and comprehensive technique for investigating non-traditional information access systems”. She concludes, among other things, that due to the complexity of user interactions with the information retrieval system, “Only by measuring each sub-task separately from the others is it possible to gain the micro-view needed to assess the effectiveness of each component”. Although our experiment was clearly different from Petrelli's, the idea of subdividing the activity into separate tasks was also beneficial, as it enabled us to carry out a more detailed analysis. As Petrelli notes, this kind of evaluation is generally used to test IR systems or confirm/refute specific hypotheses. In this scenario, useful results can only be obtained if the data extracted from the experiment is rich in both objective measures (time spent by users, number of queries and terms) and subjective ones (such as users' opinions). She also recommends analysing the data both qualitatively and quantitatively. Our experiment met these criteria, as we have used both objective and subjective measures, as well as a thorough usability study – which we think makes the results of our experiment yet more reliable.

Spink [21] created a study focused on evaluating the usability and effectiveness of the Inquirus Web search engine, “including changes users experience in their information problem and information seeking stages as a result of their interaction with Inquirus, and attempt to measure those changes”. It was based on 22 users who volunteered to fill in a questionnaire and be audiotaped while interacting with the system. We also used a questionnaire, although audiotaping and transcribing users' impressions was not practical in our situation. Nevertheless, we used a larger sample of 72 users, which we believe substantially enhances the possibility of finding trends and patterns in the data.

Van Velsen et al. [22] review the relevant literature on user-centred evaluation, and highlight the evaluation of personalised IR systems. They recommend the use of working prototypes when evaluating systems from scratch, and questionnaires are noted as among the most widely used methods for data collection. Usability, perceived usefulness and intention-to-use (which we included in our study) were the three most common variables measured in these questionnaires. There is a recent paper by Merčun, Žumer and Aalberg [23] that details two user studies which evaluated “the application of a visual presentation of bibliographic work families”; these also provide relevant results about the use of hierarchical visualisation techniques. The study was

based on the results of two experiments – a controlled one, which tested several prototype designs, and an observational one that complemented the results of the previous experiment and examined participants' interaction with the interface and level of understanding of the information analysed in the absence of "specific guidance or tasks". The first experiment evaluated several parameters such as "time, success, use of navigation features and ease of use", while the second one was focused on participants' behaviour, the quality of their answers and usability issues. Overall, Merčun, Žumer and Aalberg conclude that substantial improvements can be made by means of developing advanced visualisation and retrieval features, but these also require the library community to invest in high-quality bibliographic data and structures.

Baker et al. [24] examine "how semantic web standards and linked data principles can be used to make the valuable information assets that library create and curate [...] more visible and re-useable outside of their original library context". The goal of our experiments was precisely to measure the benefits of these kinds of systems for users in terms of retrieval and usability; these benefits are key to efforts to apply semantic technologies in library environments.

Regarding linked-data UIs, a set of tools and applications have emerged including search interfaces (e.g. Falcons), browsing interfaces (e.g. Disco, Tabulator [25], Sigma), and visualisation techniques [26]. There are also faceted browsers designed specifically for RDF data (e.g. Sparallax). Most of these studies provide generic views of resources and mechanisms to search/browse resources and their relationships. Consequently, the interfaces are capable of handling different data but the interface and user experience are not tailored towards any specific domain or tasks. More specialised UIs do exist, such as Saffron [27], or CONTENTUS [28] which combine Semantic Web technologies for representing relations between media and entities such as persons or locations with a full-text approach and exploratory search features. Datos.bne.es falls into the specialised UIs category, and explores new ways of connecting and presenting cultural content to users. Nevertheless, it takes a pragmatic approach where (1) full coverage of a large catalogue and (2) providing comparable functionalities to traditional OPAC systems are a priority.

7. Conclusion

Few semantic-based bibliographic systems and applications are accompanied by formal user-studies. This situation leaves the question of to what extent they may enhance user experiences unanswered. To mitigate this lack of empirical data we present two user-centred within-group empirical studies in this paper. To the best of our knowledge, our study is the first to compare a novel semantic-based bibliographic system with a traditional, widely deployed online access catalogue. We set out three core hypotheses, and we designed controlled experiments to evaluate our hypotheses following standard and well-established methodologies. These core hypotheses all shared the underlying assumption that semantic and linked-data technologies provide a good basis for building novel applications to satisfy users' information needs.

The results of our first empirical experiment, a task-solving scenario linked to the digital humanities and information sciences, provide both quantitative and qualitative evidence for our initial hypotheses. In particular, regarding task completion time (Hypothesis 1) and visited pages (Hypothesis 2), our results provide statistically significant evidence that semantic and linked-data technologies can have a positive impact on a typical search-and-retrieval scenario. The analysis of users' characteristics shows that age is a factor which affects time required to complete the

Table 6
Tasks in the task-solving scenario.

| Task # | Description |
|--------|--|
| 1 | Retrieve the full name, and years of birth and death of Santiago R. Cajal. |
| 2 | Retrieve the title of three works by Santiago R. Cajal. |
| 3 | Retrieve the date and place of publication of "El mundo visto a los ochenta" ("The world seen at 80 years old"), autobiographic work by Cajal. |
| 4 | Retrieve the title of three works about Cajal. |
| 5 | Retrieve the English title of the work "Los recuerdos de mi vida" ("Recollections of my life"). |
| 6 | Retrieve the title of a work about Dioscorides' "De materia medica". |
| 7 | Retrieve the year of an edition of a version in Ancient Greek of "De materia medica" which is available in digital form. |
| 8 | Retrieve the URL of a digital version of a portrait of Cajal. |
| 9 | Retrieve the name of the author of a work about the subject "Tinción" ("Staining or dyeing"). |
| 10 | Retrieve the author and the subject of the work "La ciencia española" ("The Spanish science") |

tasks. Other characteristics such as year of study, or previous experience with OPAC or datos.bne.es, did not have a significant effect. Nevertheless, our results suggest that datos.bne.es provides better results compared to OPAC, not only for specialists (e.g. librarians) but for a wider range of users. Additionally, we have provided an analysis of per-task performance and linked it to the underlying ontology and core UI features of the datos.bne.es system. Notably, our results indicate that the more complex the navigation, the more substantial the difference between the systems.

Furthermore, we also conducted a fine-grained user-satisfaction test. The results of our study show that users were more satisfied (Hypothesis 3) with the novel system.

Since these experiments were carried out, some enhancements have been implemented in the datos.bne.es system – most highlighted by the free text questions in the satisfaction questionnaire. For instance, the thematic search originally mixed topics, works, editions and other types of records. This was improved and now only the BNE subject headings are used. Subject headings are simple but normalised controlled-vocabularies that permit the indexing of works according to a predefined set of categories. This should allow for more efficient subject browsing and retrieval.

The results of the study highlight the benefits of linked-data based systems and there is certainly a strategy from the BNE to eventually make datos.bne.es the main entry-point to their catalogues. This is shown by the continuous investment in the project since 2011. The results of the current study are known to BNE staff and this is one of the motivations for their continued support for the system. However, there are other barriers (e.g. organisational) impeding a full switch to this kind of system, and the authors of this study evidently cannot represent the opinions of the BNE.

To conclude, as discussed throughout the paper, our study and results have a number of limitations, for example that the evaluation used specific hardware (personal computers) and particular research tasks (although very generic), and there is clearly more work to be done in the future. Specifically, further satisfaction tests should be conducted in order to gather more evidence and insights. Also, additional in-depth analysis of the impact of semantic modelling and UI features could help us to understand the implications of these technologies for end-user applications better, and thus lead to improved semantic solutions. Finally, our results suggest that the new system has a positive impact on users irrespective of their different backgrounds and skills, and a broader study with different user-profiles (e.g. including more of the general public) could provide stronger evidence for such an impact.

Table 7

QUIS questionnaire. The user can specify his/her agreement level selecting one value in the 1–7 range between the minimum and the maximum.

| ID | Min | Max |
|--|------------------|----------------|
| Overall reaction to the application | | |
| S-01 | terrible | wonderful |
| S-02 | difficult | easy |
| S-03 | frustrating | satisfying |
| S-04 | inadequate power | adequate power |
| S-05 | terrible | wonderful |
| Presentation | | |
| Reading text on the screen | | |
| S-P1 | hard | easy |
| Organization of messages/actions | | |
| S-P2 | confusing | very clear |
| Sequence of messages/actions | | |
| S-P3 | confusing | very clear |
| Terminology and system information | | |
| Use of terms throughout system | | |
| S-I1 | inconsistent | consistent |
| Messages are displayed in a similar position across tasks | | |
| S-I2 | never | always |
| Data are requested to the user in a similar way across tasks | | |
| S-I3 | never | always |
| The system reports the progress of actions across tasks | | |
| S-I4 | never | always |
| Error messages | | |
| S-I5 | unhelpful | helpful |
| Learning | | |
| Learning to operate the system | | |
| S-L1 | difficult | easy |
| Exploring new features by trial and error | | |
| S-L2 | difficult | easy |
| Facilities to remember names and how to perform the actions | | |
| S-L3 | few | many |
| Performing tasks is intuitive and straightforward | | |
| S-L4 | never | always |
| Help messages | | |
| S-L5 | unhelpful | helpful |
| Supplementary reference material | | |
| S-L6 | confusing | clear |
| System capabilities | | |
| System speed | | |
| S-C1 | too slow | fast enough |
| System reliability | | |
| S-C2 | unreliable | reliable |
| System tends to be ... | | |
| S-C3 | unstable | stable |
| Correcting your mistakes is | | |
| S-C4 | difficult | easy |
| Designed for all levels of users | | |
| S-C5 | never | always |

Table 8

Free text questions in the questionnaire.

| ID | Question |
|-----|---|
| Q-1 | List the most NEGATIVE aspects related to the use of the portal |
| Q-2 | List the most POSITIVE aspects related to the use of the portal |
| Q-3 | Additional comments |

Acknowledgments

This work was partially funded by projects RTC-2016-4952-7 (esTA) and TIN2016-78011-C4-4-R (Datos 4.0), from the Spanish State Investigation Agency of the MINECO and FEDER Funds.

Appendix A. Tasks in the task-solving scenario

Table 6 shows the description of the tasks given to users, translated into English.

Appendix B. Satisfaction questionnaire

Table 7 shows a text version of the QUIS questionnaire. Table 8 shows the 3 free text questions.

References

- [1] T. Heath, C. Bizer, Linked data: Evolving the web into a global data space, *Synth. Lect. Semantic Web Theory Technol.* 1 (1) (2011) 1–136.
- [2] B. Haslhofer, A. Isaac, data.europeana.eu: The europeana linked open data pilot, in: *International Conference on Dublin Core and Metadata Applications*, 2011, pp. 94–104.
- [3] A. Simon, R. Wenz, V. Michel, A. Di Mascio, Publishing bibliographic records on the web of data: opportunities for the BnF (French National Library), in: *The Semantic Web: Semantics and Big Data*, Springer, 2013, pp. 563–577.
- [4] H. Al-Kilidar, K. Cox, B. Kitchenham, The use and usefulness of the ISO/IEC 9126 quality standard, in: *Empirical Software Engineering*, 2005. 2005 International Symposium on, IEEE, 2005, pp. 126–132.
- [5] D. Vila Suero, E. Escolano Rodríguez, Linked data at the spanish national library and the application of IFLA RDFS models, *IFLA ScatNews* 1 (35) (2011) 5–6.
- [6] D. Vila-Suero, B. Villazón-Terrazas, A. Gómez-Pérez, datos.bne.es: A library linked dataset, *Semantic Web J.* 4 (3) (2013) 307–313.
- [7] D. Vila-Suero, A. Gómez-Pérez, datos.bne.es and MARiMBA: an insight into library linked data, *Library Hi Tech.* 31 (4) (2013) 575–601.
- [8] G. Charness, U. Gneezy, M.A. Kuhn, Experimental methods: Between-subject and within-subject design, *J. Econ. Behav. Organ.* 81 (1) (2012) 1–8.
- [9] P. Ingwersen, K. Järvelin, *The Turn: Integration of Information Seeking and Retrieval in Context*, Springer, 2005.
- [10] G. Cumming, F. Fidler, D.L. Vaux, Error bars in experimental biology, *J. Cell Biol.* 177 (1) (2007) 7–11.
- [11] B. Efron, R. Tibshirani, Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Statist. Sci.* 1 (1) (1986) 54–75, <http://www.jstor.org/stable/2245500>.
- [12] J.P. Chin, V.A. Diehl, K.L. Norman, Development of an instrument measuring user satisfaction of the human-computer interface, in: E. Soloway, D. Frye, S.B. Sheppard (Eds.), *Interface Evaluations. Proceedings of ACM CHI'88 Conference on Human Factors in Computing Systems*, 1988, pp. 213–218.
- [13] M. Rico, O. Corcho, V. Méndez, J.M. Gómez-Pérez, A semantically enhanced UPnP control point for sharing multimedia content, *IEEE Internet Comput.* 15 (6) (2011) 58–64.
- [14] M. Rico, J. Ramírez, D. Riofrío-Luzcando, M. Berrocal-Lobo, A cost-effective approach for procedural training in virtual worlds, *J. UCS* 23 (2) (2017) 208–232.
- [15] P. Borlund, P. Ingwersen, The development of a method for the evaluation of interactive information retrieval systems, *J. Doc.* 53 (3) (1997) 225–250.
- [16] P. Borlund, Experimental components for the evaluation of interactive information retrieval systems, *J. Doc.* 56 (1) (2000) 71–90.
- [17] P. Borlund, The IIR evaluation model: a framework for evaluation of interactive information retrieval systems, *Inform. Res.* 8 (3) (2003).
- [18] B. Ricardo, R. Berthier, *Modern Information Retrieval: The Concepts and Technology Behind Search*, second ed., Addison Wesley, 2011.
- [19] A. Díaz, A. García, P. Gervás, User-centred versus system-centred evaluation of a personalization system, *Inform. Process. Manag.* 44 (3) (2008) 1293–1307.
- [20] D. Petrelli, On the role of user-centred evaluation in the advancement of interactive information retrieval, *Inform. Process. Manag.* 44 (1) (2008) 22–38.
- [21] A. Spink, A user-centered approach to evaluating human interaction with web search engines: an exploratory study, *Inform. Process. Manag.* 38 (3) (2002) 401–426.
- [22] L. Van Velsen, T. Van Der Geest, R. Klaassen, M. Steehouder, User-centered evaluation of adaptive and adaptable systems: a literature review, *Knowl. Eng. Rev.* 23 (03) (2008) 261–281.
- [23] T. Merčun, M. Žumer, T. Aalberg, Presenting bibliographic families using information visualization: Evaluation of FRBR-based prototype and hierarchical visualizations, *J. Assoc. Inform. Sci. Technol.* 68 (2) (2017) 392–411.
- [24] T. Baker, E. Bermès, K. Coyle, G. Dunsire, A. Isaac, P. Murray, M. Panzer, J. Schneider, R. Singer, E. Summers, W. Waites, J. Young, M. Zeng, Library linked data incubator group: final report, *Tech. Rep.*, World Wide Web Consortium (W3C), 2011.

- [25] T. Berners-Lee, J. Hollenbach, K. Lu, J. Presbrey, E. Prud'hommeaux, M.M.C. Schraefel, Tabulator redux: browsing and writing linked data, in: C. Bizer, T. Heath, K. Idehen, T. Berners-Lee (Eds.), LDOW, in: CEUR Workshop Proceedings, vol. 369, CEUR-WS.org, 2008.
- [26] C. Christou, C. Angus, C. Loscos, A. Dettori, M. Roussou, A versatile large-scale multimodal vr system for cultural heritage visualization, in: Proceedings of the ACM Symposium on Virtual Reality Software and Technology, ACM, 2006, pp. 133–140.
- [27] F. Monaghan, G. Bordea, K. Samp, P. Buitelaar, Exploring your research: sprinkling some saffron on semantic web dog food, in: Semantic Web Challenge At the International Semantic Web Conference, vol. 117, 2010, pp. 420–435.
- [28] J. Nandzik, B. Litz, N. Flores-Herr, A. Löhden, I. Konya, D. Baum, A. Bergholz, D. Schönfuß, C. Fey, J. Osterhoff, J. Waitelonis, H. Sack, R. Köhler, P. Ndjiki-Nya, CONTENTUS—technologies for next generation multimedia libraries, *Multimedia Tools Appl.* 63 (2) (2013) 287–329.