# BibShare: An Interoperable System to Access and Maintain Bibliographic References

José H. Canós[1] and Eduardo Mena[2]

[1] Department Sistemes Informatics i Computacio
University Politecnica de Valencia
Spain
jhcanos@dsic.upv.es
[2] Department IIS, University of Zaragoza,
Spain
emena@posta.unizar.es

**Abstract.** The management of bibliographic data that can be cited from scientific documents is a very important issue for authors of publications. Unfortunately, the storage, update and bibliography generation can become a time consuming task for researching groups. Fixing inconsistent, obsolete or incompleted references is done individually by each author or group of authors, although people working on similar issues store similar bibliographic references.
Most of these problems can be avoided if bibliographic data is maintained by third parties and authors simply access such data in order to automatically build the bibliography of their publications. However, although several attempts exist in dealing with significant collections of bibliographic data, there is a lack of tools that allow users of different word processors to access such collections.
In this paper we introduce BibShare, a system based on a federation model that is able to provide authors and collection managers with mechanisms to define, maintain and cite bibliographic references.

## 1 Introduction

The quality and/or usefulness of any information source (either a book, a technical report, a scientific paper, a thesis, etc) is often measured in terms of both the accuracy of its contents and the list of references to previous related work that it includes. For a writer, especially in the academic/scientific world, the task of keeping an up-to-date, well-organized bibliography database is becoming a key point on his/her daily work. However, good reference management support has been for years one of the main lacks of word processing systems and several solutions have been proposed.

Bibliography managers are software systems allowing their users to insert cites into specific places of a document and generate on demand the bibliography list in the appropriate place of the document. For instance, for users of Microsoft Word [5], software firms have developed tools like ProCite [9] and End Note [8],

that are widely used by writers, especially in specific domains for which there exist good collections of both references and bibliographic styles. Even some scientific digital libraries offer to their users the bibliographic record of the papers in such formats (e.g. Hihgwire [7]). Other tools offer similar features, but an exhaustive review of them is out of the scope of this paper; we refer the reader to Google's catalog of bibliographic utilities [2]. These systems offer facilities to organize bibliography collections, as well as database-like services to retrieve references from a collection. However, their main drawback is the proprietary format in which records are created and stored, and therefore such bibliographic data cannot be used by other word processors without a high cost.

In the case of TEX [12], the typesetting system widely used in Computer Science and other fields, the BibTEX tool supports bibliography management based in the very popular format of `.bib` files. They are plain text files in which bibliographic records are collections of tagged strings following a predefined metadata set. Internet is plenty of bibliographic records in `.bib` format, and the most important bibliographic collections in mathematics, physics, and computer science, just to name a few, provide users with the corresponding BibTEX records of the papers they contain. As a drawback, we must mention the lack of support for collection management that makes difficult to retrieve records, especially in large `.bib` archives. These problems are derived from the fact that TEX users have converted the `.bib` format, which is the source code for BibTex, into the standard format for storing references.

All the mentioned systems share the idea of storing bibliographic data locally, according to which they are usable as far as a user or group of users create and maintain a citation database, from which the references are taken at the bibliography generation time. This has produced a large amount of duplicate information, as many of the key references in a given area are very likely to be present in every collection of citations. Moreover, the cost in time of maintaining such personal/group collections is fairly high.

An even more important problem is that of users that occasionally write documents using sometimes TEX and sometimes other word processors (typically Microsoft Word): if one wants to use his/her reference collection in both systems, he/she must maintain two different databases with the very same contents stored in different formats. This is obviously not desirable but more frequent than expected. Moreover, as `.bib` files are plain text files, handling large collections of citations is rather tedious.

## 2 BibWord

In 1997, one of the authors (Canós) started a project called BibWord with a double goal: on the one hand, to create a free bibliography manager for Word users; on the other hand, to ease the reuse of available BibTEX collections by creating "`.bib` loaders". Several versions of BibWord have been released since then. The very first one was usable only in local mode as the access to the reference database was made via ODBC drivers. Soon it was enhanced with a

client/server version in which Microsoft Word clients could access the reference database via TCP/IP calls. Additionally, a web-based access was developed for reference collection management. A description of this version was published in [1].

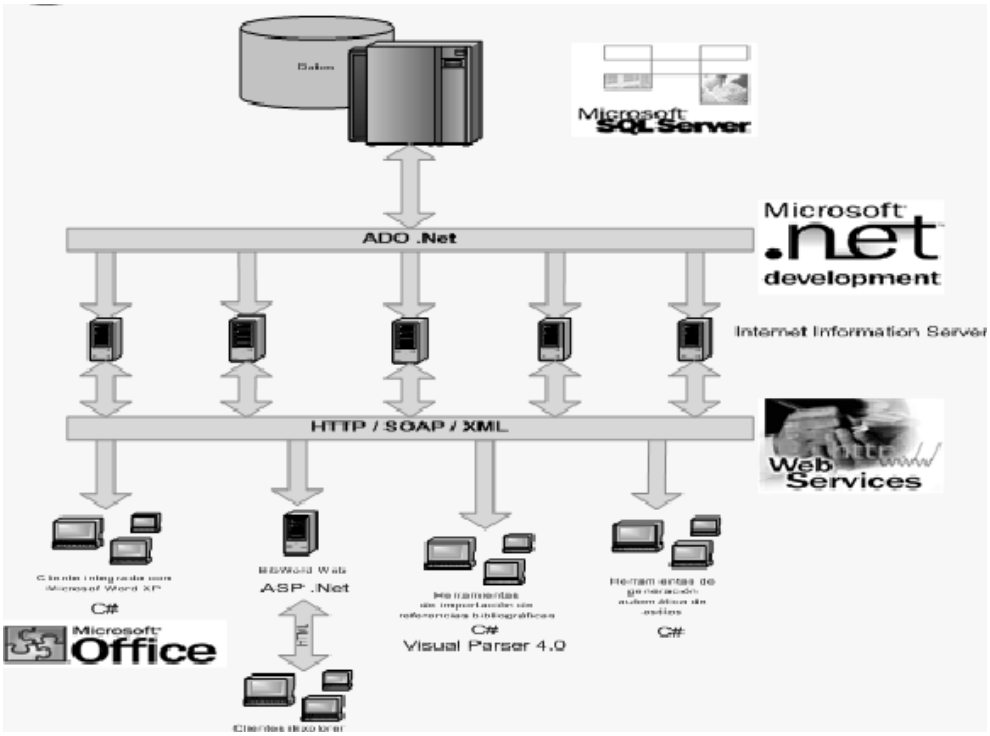


**Fig. 1.** Architecture of BibWord 2.0

Since the publication of this work, more than 1,600 users have downloaded BibWord. Many of them did it because they wanted to use their BibTEX collections in Word. Now we are about to release a brand new version we have implemented using the Microsoft .Net platform and web services (see figure 1). At the core of BibWord v2.0 there is a reference collection, which is stored in a reference database whose schema is compliant with the BibTEX schema. The access to the database is provided as a set of web services that are invoked by different clients:

- A web interface (BibWordWeb) for creating, deleting, updating, and retrieving references.
- Tools for converting references from other formats into the BibWord internal format. Specifically, a `.bib` file converter has been developed.

– For each word processing system, a client that inserts cites into the text and automatically generates the bibliography list according to some bibliography style.

Currently, the bibliography manager for Word is available, and we plan to develop managers for other systems, including Word Perfect and LaTeX.

But even with this improved version, the problem of reference harvesting and management remains unsolved. That is why we are trying to change from a bibliography-owning philosophy to a bibliography-sharing one, in which large reference repositories can be accessed (remotely) by word processors. This would lead to BibShare, a system basically similar to BibWord, but in which the BibWord database is replaced by existing citation databases like DBLP [4], CiteSeer [3], or Highwire [7], to name a few of the most relevant ones. The federated system resulting from the integration of these databases will be accessible in such a way that queries to the interface will be transformed into local queries to specifically-designed APIs in each collection.

## 3 A Federation Model for Bibliography Management

The other author of this paper (Mena), has been working on a system to harvest bibliographic references from different data repositories [11, 10]. The proposed system takes advantage of mobile agents [6] to achieve its goal. But the important point is that, in their approach, bibliographic data must be seen from a different point of view: instead of creating (and maintaining) many bibliographic repositories, which leads to problems concerning inconsistency, accesibility, etc, they advocate recollecting data from each node and storing them into a global consistent node, that provides users with tools to access such data. Thus, instead of using the tools provided by query processors to build bibliography (if any), query processors should access to that bibliography repository with the help of corresponding mediators. The system developed was inspired in type of bibliographic data management proposed by LaTeX.

We plan to apply the very same philosophy to the development of BibShare. It will allow access to Citation collections through specifically designed web services, which will be invoked in the very same way by the bibliography manager(s) of the word processor(s), regardless what specific word processing system we are using at a given moment. Moreover, by having these databases available, the duplication of citations in individual reference collections is avoided.

## 4 Conclusions

Managing bibliographic data is one of the most time consuming task in the writting of publications, mainly when different word processors are used. In this paper we have introduced an approach that advocate developing tools to access huge collections of bibliographic data, like Citeseer and DBLP. This is a a real

need of anybody involved in modern research, and we believe that BibShare can become a key tool for writers.

As future work we plan to integrate our complementary initial proposals on this issue to develop a system that implements this philosophy of work. In addition, we are contacting with the managers of important bibliographic data repositories in order to have these systems open for bibliographic queries in our future environment.

## References


1. J.H. Canós. A bibliography manager fot microsoft word. *ACM Crossroads, Special Issue on Windows programming. Summer 2000*, 2000. http://www.acm.org/crossroads/xrds6-4/bibword.html.
2. Google. Google, catalog of bibliographic utilities, 2002. http://directory.google.com/Top/Reference/Libraries/Library_and_Information-_Science/Technical_Services/Cataloguing/Bibliographic_Utilities/.
3. NEC Research Institute. CiteSeer, Scientific Literature Digital Library, 2002. http://citeseer.com.
4. Michael Ley. DBLP: Computer Science Bibliography, 2002. http://www.informatik.uni-trier.de/
5. Microsoft. Microsoft word, 2002. http://www.microsoft.com/office/word/default.asp.
6. S. Papastavrou, G. Samaras, and E. Pitoura. Mobile agents for WWW distributed database access. *IEEE Transactions on Knowledge and Data Engineering*, 12(5):802–820, 2000.
7. Highwire Press. Highwire, Library of the Sciences and Medicine, 2002. http://highwire.stanford.edu.
8. ISI Researchsoft. EndNote, 2002. http://www.endnote.com.
9. ISI Researchsoft. ProCite, 2002. http://www.procite.com.
10. J.A. Royo and E. Mena. Uso de agentes móviles para la búsqueda y recuperación de información bibliográfica. *Revista Interamericana de Nuevas Tecnologías de la Información, ISSN 0122-3356*, 6(4):52–63, October–December 2001.
11. J.A. Royo and E. Mena. Gestión de bibliotecas digitales de publicaciones de investigación. In *III Jornadas de Bibliotecas Digitales (JBIDI'2002), Madrid, Spain*, November 2002.
12. TEX Users Group (TUG). Tex, 2002. http://www.tug.com.