semantics

aggregation

knowledge

posts

topics

comments

preservation

tags

Intelligent blog preservation

Potential users users of BlogForever fall into the following categories:

- Memory Institutions: Future researchers will not be able to study our lives without the internet. If we don't preserve the web we will have a significant gap in our knowledge of our history. Libraries, archives, museums have recognised the need to preserve internet records for future generations.

- Universities, and Information Centres: Educational organizations which would like to use certain groups of blogs based on their characteristics. These organizations could use certain content from BlogForever for educational or informational purposes for their users.

- Researchers: Academics who would like to aggregate knowledge and information from blogs for a variety of reasons such as scientific research.

- Blogging communities: Groups of bloggers will be able to preserve their blogs effectively and reliably using a turn-key solution.

- Businesses: Businesses which would benefit from monitoring the blogosphere for specific trends in business and knowledge domains.

- Search engines will be able to obtain transparent access to BlogForever's index using any of the supported protocols such as Z39.50 and SRU.

Partners

- Aristotle University of Thessaloniki
- European Organization for Nuclear Research (CERN)
- University of Glasgow
- University of Warwick
- University of London Computer Centre
- Technische Universitŕt Berlin
- CyberWatcher
- Sofware Research and Development and Consultancy Ltd.
- Tero Ltd.
- Populis
- Phaistos Networks S.A.
- Altec Research S.A.

Contact

blog forever

A software platform capable of harvesting, preserving, managing and disseminating blogs.

- BlogForever develops robust digital preservation, management and dissemination facilities for weblogs.

- These facilities capture the dynamic and continuously evolving nature of weblogs, their network and social structure, and the exchange of concepts and ideas that they foster; pieces of information omitted by current Web Archiving methods and solutions.

Scientific Objectives

- Study weblog structure and semantics

- Define a robust digital preservation policy for weblogs

- Implement a weblog digital repository

- Implement specific case studies

Advances to the state of the art

- Definition of a generic data model for weblog metadata and semantics

- Weblog digital preservation strategies

- Weblog spider

- Weblog digital repository web application

BlogForever creates a sofware platform capable of aggregating, preserving, managing and disseminating blogs.

Blogs

Unstructured information →

Web services, Blog APIs →

blog forever
Weblog Spider

- HTML data extraction engine
- Web services extraction engine
- Real-time weblog monitor
- Spam and noise filtering

{ **XML Metadata**

Weblog posts
Comments
Metadata (Content, DRM, Technical)

Semantic information
Microformats
Associations between different blogs, posts, users, etc. }

Companies

Libraries, Information centres

Research

Web services
Web interface

- Digital preservation
- Data migration
- Quality assurance checking
- Collections curation
- Public access APIs

- Personalized services
- Data mining, memetracking
- Information retrieval
- Public Web Interface, Browse, search, export

blog forever
Weblog Digital Repository

Impact

- The final output of BlogForever is a simple weblog digital archiving solution that any user, user group or institution could use to preserve their weblog(s) and ensure their authenticity, integrity, completeness, usability, and long term accessibility as a valuable cultural, social, and intellectual resource.

- Usage can be approached in two different ways:
  SaaS/Cloud service: online service to create focused repositories/collections of blog archives.
  Downloadable software: can be installed by organizations which need to create their own blog preservation workflow by deploying their own blog archives.