Enabling Information Literacy despite Information Chaos

No matter how one defines information literacy, it always boils down to the abilityof finding relevant information and doing something intelligent with it.

Abstract

The ease and low cost of distributing content on the Internet today enable an overwhelming mass of (often unstructured) information to build up data smog, making it very difficult to isolate crucial information and new knowledge for science and business.

Documents must be search friendly with meaningful embedded or attached metadata for easy discovery and auto- classification in the users' knowledge base so that the information can rapidly be found again when needed for work. Documents must also be user friendly and include navigation to allow quick access to specific information within long documents.

This requires some discipline on the part of authors and publishers to produce suitably formatted or "well-behaved" documents[1], i.e. documents that are both search friendly and user friendly.

Users, on the other hand, must have suitable software available that can read metadata, index and organize incoming information for a better overview and facilitate finding it again in the mass of material that accumulates over time. Such software must be able to handle all kinds of data including books, training material, reference works (typically PDF documents, images, videos and podcasts) as well as remote links to such material.

Information chaos is pre-programmed

Everybody masters the minimal skill of entering word(s) in a search engine and then sees thousands of links exposed. To know which links provide the best answers to solve the problem at hand or to provide new insights is, however, not that easy. For this reason people tend to tag or download more than one document for later off-line study.

In addition there are lots of links and downloadable information served via email lists and social media channels which may not be relevant for the moment, but are kept anyhow since they could perhaps be useful in the future. Case studies, guidelines, and reference works are typical examples.

The result is known as information chaos which consists of an overcrowded bookmark list and a hard disk or cloud drive full of forgotten documents.

Scholars who have the wherewithal to find citation relevant metadata in repositories and through metadata harvesters, also find themselves confronted with this phaenomenon. Open Access, the academic pressure to publish and to cite and the ease of self-publishing contribute to an exponential growth of the body of (scientific) knowledge on the Internet which becomes daily a bit less transparent.

As individuals we are also multitasking[2], mixing digital life and digital work. Stumbling upon an interesting video while reading ones email or searching for a topic and wanting to save it for later viewing at home is multitasking. Doing this a few times per week also produces content overload which makes it difficult to isolate new knowledge for the purpose of learning and gaining new understanding.

The challenge

Keeping track of all these items becomes a real challenge and users must learn how to manage their personal knowledge base in order to keep an overview and to quickly find relevant information when needed.

It is further important that found documents are easy to read and to peruse to save readers from having to leaf through hundreds of pages to arrive at a specific piece of information.

The key to a better organization of random information is useful metadata. Every piece of incoming data should be literacy enabled and have embedded or attached metadata, which will allow for automatic classification thereof so that the information can be rapidly located again when needed for work. The need to be able to manage knowledge increases with each download or saving in the cloud or bookmarking a web link.

The solution

Metadata

Key to automatic classification is meaningful metadata. Every document and every piece of data must be **search friendly** and have embedded or attached metadata.

Easy Navigation

Documents should be **user friendly**, i.e. documents with more than just a couple of pages should have bookmarks and, where applicable, an interactive table of contents, links to external resources and literary citations. They should not be in a proprietary format (requiring costly software for reading) or password protected, but allow text extraction via copy/paste. Their layout and font selection should be optimised for easy on-screen reading.

Engagement

Authors, those who buy authoring work and publishers must recognize the need to make all documents they offer for sharing or on the web user- and search-friendly and must be aware of the benefits thereof such as

- search engine optimization – information is easier to discover
- information inside longer documents is more readily accessible
- knowledge can effortlessly be managed at users' end
- better authors' exposure
- higher chance of being cited or referenced

- facilitates (academic) exchange of data

- they must also know that making documents interactive and embedding metadata does not necessarily require any extra work if properly planned and some simple rules (such as the consistent use of styles) are observed.

An author who has worked a month on a paper can certainly spend 10 more minutes to write down some keywords plus a description, the typesetter who produces the table of contents anyhow has only to check a single box before exporting to PDF and the publisher can easily import an XMP file containing metadata into the final document to make it "well-behaved".

Software

Users need access to intelligent and free software tools that can automatically classify incoming data sources based on embedded or attached metadata and thereby enabling them to better organize their personal knowledge base.

This applies in particular to users who must collect lots of reference and instruction material for their work, but also to students, authors and scholars who need to keep track of and share knowledge:

One example of such software is the free digi-libris Reader[3] which includes some unique features not found in other organization tools.

- digi-libris Reader is very easy to use, just drag a document or copy a web link into the main window – it will automatically be classified and alphabetically indexed

- available for Windows, Mac and Linux operating systems with English, French, German, Russian and Spanish menu language and metadata attribute names

- can mix documents, images, videos, any other type of data as well as remote links and even physical objects and display side by side in a single list for overview

- reads and permits editing and adding metadata in various standards or user defined attribute/value pairs for cross-disciplinary usage

- easy citation generation using generic styles or CSL styles sheets based on available metadata

- metadata pasting from unstructured documents if no metadata is available

- data sharing via dMeta, a file format to bundle any type of data with its descriptive metadata as XMP sidecar file – especially useful where metadata cannot directly be embedded such as in scanned documents, work in progress, presentations or data generated by an instrument

- content providers and educational establishments may offer digi-libris Reader free of charge to their readers, students or members and even include their own topics and keywords lists to facilitate the exchange of knowledge.

Conclusion

Literacy enabled information is information that is easy to discover and easy to peruse. All downloadable documents should be "well-behaved", i.e. easy to navigate with bookmarks and, where applicable, with an interactive table of contents and contain embedded or attached metadata for automatic classification and sharing of knowledge (given suitable software).

Information consumers should clean up their filing system and organize their personal knowledge base using intelligent tools in order to gain better insights and learn more and faster.

Call for action

Institutions distributing instructional material should make sure that their members are optimally enabled to profit from such material by distributing only well-behaved documents and equipping them with adequate tools.

Educational establishments should make sure that students learn, as early as possible, to understand the value of metatada and how these can best be used to organize knowledge and generate bibliographic references.

All authors and content providers should use their know-how and influence to insist that publishers **make their offerings literacy enabled** and that software producers include automatic meta-data parsing in their programs.

---

[1]   This PDF document is a typical example of a well-behaved document. It has bookmarks and, above all, meaningful embedded metadata including a thumbnail image in addition to a telling file name and thus would classify itself automatically in a users' knowledge base (given suitable software).

[2]   Multitasking is the most frequently played Internet game designed to deliver total confusion (Miescher 2015) and it doesn't even exist, it is simply switching between doing two things badly (Huffington 2015).

[3]   Please view the YouTube demo and
        visit https://digi-libris.com for more details
        or to download the program for testing.

        Institutions wanting to offer the program free of charge to their constituents, can either provide a link to the download page or request a bespoke version which could include a custom designed front page and personalized topics- and keyword-lists.

Geneva, September 26, 2015