



ATG Press Research Kit

Inside Document Management I: Document Creation

March 2006

A research document from the
Aria Technology Group Science & Technology Office

Important Note: The information, which is provided by this document, is devoted to the entire world for improving creativity, and therefore developing new job opportunities around the globe, in the field of information technology. So, no part of the proposed information is protected by any intellectual property right. However, the document itself, is copyrighted by Aria Technology Group and is a priced material. No part of this document could be published, in any form of media, without written permission of Aria Technology Group.

Copyright © 2006 by Aria Technology Group
<http://www.ariatg.com/enterprise>



OVERVIEW

The growing of the Internet in recent years has attracted many attentions to the concept of document. The reason is very clear: everyday millions of documents are updated or added to the net (the so-called information explosion) and the management of these activities would cost billions of dollars worldwide. In addition to this, new and complex technical issues should involve in this huge-scale management. As the term “document” is a very general concept, we believe any effort for giving a precise definition to it would come to an apparent limitation to the field, instead of opening a new way. Traditionally, document is a record of some activities. Technically, document is a package of information. Intuitively, document is the heart of information technology. In the current IT world, the most important issue regarding document is its manageability and other efforts such as document definition/standardization are just used as leverage in this issue. Addressing the issue, we attempt to show a new perspective to the critical problems in document management. To this end, we divide the document management domain into two main parts: document creation and document presentation. In this text, we focus on the first part and try to briefly describe the problems involve in the realm of document creation including why, how, and by whom a document can be created. Detailed solutions to the mentioned problems are expected to appear in a series of our future research kits, but in this research kit, we only introduce the concept behind of those solutions.

1. WHY TO CREATE

At the first look, the question of why a document should be created may seem a very general and useless question. However, it is not possible to manage a document without knowing the intention or motivation of its creation. The current trend in the IT world is to find the most referred and popular roles of documents and suppose them as the major intentions of document creation. It is an apparent bottom-up approach to the problem and as any other bottom-up approach would limit our viewpoint to the current and local traditions. For example, business transactions are one of the most referred forms of documents at present. Therefore, for many people in and out of the IT industry, document management is equivalent to a new way of business management and document creation is a basic business activity.

The most important problem with the bottom-up approach is that any change of behavior in document usage would impose some new adaptations to the document management methods. These adaptations are both costly and technically problematic. Instead, if we find a top-down approach to the problem, our management methods could be robust enough to tolerate many behavioral changes. The reason is that in any top-down approach, the essence of the problem is investigated, regardless of its current applications. However, the top-down approach has a major drawback, i.e., abstraction. To prevent it, we should take into account all of the historically proposed roles of documents regardless of their popularity.

Interestingly, the top-down approach can result a brief but comprehensive categorization for the different intentions of document creation. The following is the proposed basic categories:

1. Addressing some activities or events beyond (before/after) their occurrence time (family photograph, shopping list)
2. Addressing some activities or events beyond their occurrence place (Internet voting, live broadcasting)
3. Addressing some activities or events beyond their occurrence time & place (music record, newspaper)

The above categories are basic, i.e., they do not include sub-categories. In these categories, the term “activity” means everything that can be done by a human being including art and imaginative works, but the term “event” means something real that can happen in the physical world. In the front of each category,



two simple examples are cited to introduce the concept better. These categories are based on the fact that we are living in a world that can be addressed by time-space in its physical aspects. Although the cyberspace can leap up some boundaries in time and space, but it cannot change their essence.

2. HOW TO CREATE

Intuitively, any document needs a medium to be existed in the physical world, and the previous section shows this fact obviously. Therefore, the problem of how we should create a document has a direct relationship with the medium we use for it. In addition to this, it is possible to use different logical methods for creating a document using a same medium. Consequently, the problem of how we should create a document has two different aspects: physical and logical lines.

As in the current IT world most of the physical document formats are easily convertible to each other, the medium of a document does not matter any more. Rather, the logical document format has a critical role in document management. Traditionally, a logical format is proposed by some people to the industry and then it gets attentions in a rate proportional to its applicability and extendibility.

If we step over the concept of logical format, we reach the concept of logical process. This concept shows what we should do to cast information into a supposed logical document format, and in reverse, what we should do to extract the information from it. Now an important question is raised: which one of the logical format and logical process depends on the other? For the bottom-up followers, the key issue is the logical format and logical process depends on it, but for the top-down proponents the situation is reverse.

As with the previous section and for the same reasons, we prefer top-down methodology. So, the question of how we should create a document is reduced to the question of which logical process can be used to create a document. To answer this question, we should consider different logical processes, which are used in practice for document creation. This consideration is regardless of the logical format that a process deals with. The following basic categories show the possible processes:

1. Direct casting (audio/video recording, pulse recording)
2. Descriptive casting (text composing, XML file composing)
3. Computational casting (graph generating, signal composing)

In the current document management methodologies, the second category is the center of attention. The reason is that the underlying logical formats of this logical process are extensive and extendable. For many people, the extendibility means more management power. However, this extensibility has another face, i.e., more confusion. For example, if you look at the number of XML vocabularies existing around for the same thing, you will find out what we mean by confusion.

It seems the secret of true management power is the uniform coverage of all categories and aspects. Investing in only one category cannot be directed into a consistent management methodology in long term.

3. WHO CAN CREATE

In the IT world, machines (electronic devices) have a pivotal role in both of document creation and presentation. Therefore, one of the important issues in document management is the arrangement of works between human and machine. This arrangement needs a complete understanding of human's and machine's capabilities in document creation and presentation. The question of who can create a document wants to address this issue.

Before answering this question, the concept of human-oriented and machine-oriented documents should be briefly explained. A human (or machine) oriented document is one that is directly usable by a human (or machine), regardless of its creator (or presenter) that can be either a human or a machine. For example, a



computer program that is created by a human being is a machine-oriented document, and a pie-graph that is created by a computer is a human-oriented document. The electronic version of a newspaper is a human-oriented document that is created by human but is presented by machine.

Naturally, machines are the masters of direct and computational document creation, and humans are the masters of descriptive. In addition, humans and machines can collaborate with each other to create a document. The following shows the different categories of the document creators:

1. Human: document content is completely made by human (journalist)
2. Machine: document content is completely made by machine (CAD system)
3. Hybrid: document content is made by human & machine interaction (Semantic Web)

At the present time, the abilities of the first & second groups of creators are adequately known for the IT insiders, but many things should be known about the potentials of the third group. The issues such as the Semantic Web or the Globe Server are to exploit these potentials in the near future. In fact, with the worldwide growth of information, there will be no way but to make on-demand documents for users.

4. CONCLUSION

The classical management concept tells how we should allocate our *resources* to meet our *demands*. When our resource is the dispersed information and our demand is the knowledge, which is exactly the case of document management, this classical definition is hard to adopt. It is due to the intangible nature of both the information and knowledge. This fact shows why the bottom-up approach is a popular methodology in document management. In this approach, we do not need to strictly determine our resources and demands, but the instinctive usages of documents are gradually modernized.

In this research kit, we tried to open the door to the usage of classical management models in document management. To this end, we used a tricky way to detect the resources and demands: instead of digging them out directly, we tried to scrutinize and categorize different behaviors in document creation and presentation. These behaviors (why, how, who) can apparently show the underlying resources and demands in document management.

Now, it is possible to employ different management models in document management and compare their efficiencies regardless of the issues such as document standards and models.

BIBLIOGRAPHY

- Berners-Lee, T., J. Hendler, and O. Lassila. 2001. "The Semantic Web," *Scientific American*, May 17.
- Glushko, Robert J., Tim McGrath. 2005. *Document Engineering: Analyzing and Designing Documents for Business Informatics and Web Services*, MIT Press.
- Laleh, Farnad. 2004. "[Globe Server](#)," *ATG Press Appetizer*, Vol. 1, pp. 7-12.
- Laleh, Farnad. 2004. "[Globe Client](#)," *ATG Press Appetizer*, Vol. 1, pp. 13-17.
- Lohr, Steve. 2006. "Push to Create Standards for Documents," *New York Times*, March 3.