

Taxonomies and information architecture

Alan Gilchrist

Cura Consortium and TFPL Ltd.

0.1. Resumen

Se postula que las taxonomías difieren de manera tan fundamental de las clasificaciones y tesauros que requieren una dimensión gerencial además de la meramente técnica para su diseño, construcción e implementación. Esta aproximación sinóptica, que toma en consideración desde una perspectiva más detallada la estructura organizacional, los procesos, las herramientas, los recursos humanos y la interacción entre ellos, se denomina Arquitectura Informacional. Dicho término es definido y ejemplificado a través de la discusión de los cuatro componentes genéricos mencionados, especialmente en la medida en que afectan a la usabilidad y recuperabilidad del interfaz persona-ordenador.

Palabras clave: Taxonomías. Arquitecturas informacionales. Gestión.

0.2. Abstract

It is argued that taxonomies differ in such fundamental ways from classifications and thesauri that the approach to their design, construction and implementation require a managerial dimension in addition to the technical. This synoptic approach, which takes into account a more detailed consideration of organisational structure, processes, tools and people; and the interactions between them is called Information Architecture. This term is defined and exemplified by discussion of these four generic components as they relate to usability and findability at the computer/human interface.

Keywords: Taxonomies. Information architectures. Management.

1. Introduction

Interest in taxonomies continues to grow, both in the private and public sectors. A report from the consulting firm Delphi has predicted in a survey that some 90% of firms in a Delphi market survey (conducted at the end of 2001) were expecting to have a taxonomy strategy in place within the next 24 months (Delphi Group, 2002b). And yet, there is still confusion in some quarters as to what exac-

tly a taxonomy is. Some still regard it as another name for either a classification or a thesaurus. While taxonomies — and several types may be discerned in the literature (Gilchrist, in press)— have something in common with both classifications and thesauri, and indeed may use some of their features, there are some fundamental differences that raise the level of their consideration, construction and application to a higher level. These differences include:

- that they are enterprise-wide in a real sense, in that they are designed to transcend professional and functional boundaries, typically being mounted on corporate portals,
- that they concentrate on internally generated information (typically accounting for some 80% of what is stored by an organisation), highlighting the problem of dealing with organisational terminology,
- that information retrieval is increasingly embedded in wider applications, rather than being offered as a stand-alone and optional extra,
- that they may have to relate to, even correlate, a range of other terminologies used for other purposes within the organisation, and...
- that taxonomies are user-focused, employing the terminology and concept structures familiar to them, rather than being imposed by external information specialists.

1. Information architecture

These factors call for a more synoptic approach that, in common with others, we at TFPL, call ‘Information architecture’. Our definition is very simple: “A coherent set of strategies and plans for information access and delivery within organisations”.

Others take a wider view. Figure 1 presents the scope of Information architecture as seen by Dillon (2000) at an ASIS Conference on the theme in 2000. A new organisation, called the Asilomar Institute for Information Architecture (Asilomar is Spanish for “refuge from the sea”) (<http://www.aifia.org>) defines Information architecture as:

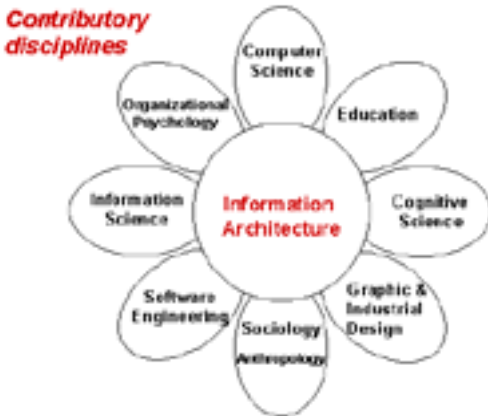


Fig. 1. The possible scope of Information Architecture (Dillon, 2000)

1. The structural design of shared information environments.
2. The art and science of organising and labelling web sites, intranets, online communities and software to support usability and findability.
3. An emerging community of practice focused on bringing principles of design and architecture to the digital landscape.

The first of these definitions is a concise working definition used for the marketing of consultancy services; the second is academic, and may form the basis for a university course; while the third is an opening offer from a new institution. What they all share is a belief in the equal importance of usability and findability at the user interface, and that the issue is strategic. They also all believe, implicitly or explicitly, that information architecture is a multidisciplinary field.

2. An Information Architecture framework

Figure 2 shows what is known as the “Leavitt diamond” (Wiggins, 2000), a simple illustration of the interlocking of the major components of any organisation.

‘Structure’ includes the more or less permanent framework of an organisation: the objectives, rules and standards that govern the arrangement of its processes. Processes includes all those activities that are designed to transform inputs into planned and controlled outputs — which are undertaken by the people of the organisation, using all the tools (hardware, software, taxonomies etc) supplied by the organisation for those purposes. The thesis behind this synoptic view is that every non-trivial innovation contemplated

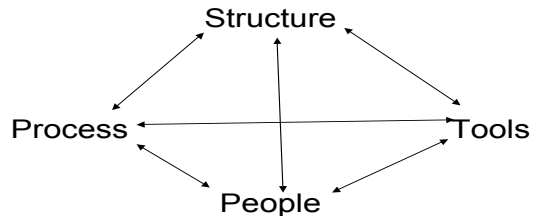


Fig. 2. The “Leavitt diamond” (Wiggins, 2000)

by an organisation must take into account the potential contributions from, and effects on, all the sub-elements within each quarter of the diamond — and above all the interactions between them, and the way in which the whole organisation will undergo change following the introduction of the innovation.

Remembering that all four quarters interact, it can still be useful to look at each of the pairs separately.

2.1. Organisation/People

This is, perhaps, the most difficult area of consideration involving, as it does, the whole range of human factors, including individual characteristics and aspira-

tions and the way in which individuals relate to groups and whole organisations. This area involves consideration of organization theory and group psychology and, as a specific phenomenon, 'information politics' about which Davenport *et al.* have written in a light-hearted but serious vein (Davenport, 1992). It is also, of course, the domain of Knowledge Management (Kelleher, 2001), which is primarily concerned with the human issues of the sharing and elicitation of tacit knowledge, and secondarily with the processes and tools that support these activities. It is at this second level that information architecture and taxonomy creation are deployed, but effectively only if within an understanding of the primary level.

2.2. Organisation/Processes

Two important features under this heading are Business Processes and the Information Life Cycle. It is vital in information architecture to understand the business processes of the organisation, so that information support may be fully integrated with workflow. One of the key business processes is the information life cycle itself, consisting principally of acquisition or generation, authorisation, indexing and storage, retrieval, communication and use (Wiggins, 2000). If business processes are not already documented, then the information architect must chart them in sufficient detail to support any information system implementations being planned.

It may also be possible to relate particular business processes to elements of the associated information life cycle. This has always been obvious in scientific research, where searching for information at the beginning of a project normally requires wide recall and correspondingly low precision, which measures reverse their values as the project develops. Some organisations are beginning to extend this principle, recognizing that information processes and requirements will change as the professional works through a task, for example a lawyer undertaking a case from brief to conclusion.

At a more generic level, it should be appreciated that different enterprise-wide applications have quite different characteristics. Figure 3 maps the different stages of the information life cycle against important features of each and a selection of final applications. Increasingly, information retrieval facilities are embedded in these corporate schemes, and the problem of reconciling and integrating these schemes begins to emerge, again with strong implications for information architecture and taxonomy building.

2.3. Organisation/Tools

The organisation must specify, procure and install tools to support business processes in a way that the users find acceptable, implying the need for their participation in the process from conception to implementation and training. Many vendors claim, and customers admit, that specification of requirements is a weak

Information Architecture – A Roadmap

Preparation/ Activities

- 1 Determine business objectives
- 2 Information / Knowledge Audit
- 3 Organise / Apply standards

Operational activities: **Black** Internal actions – **Blue** External or brought in elements – **Red**

D= Document, P=Person, B=Business, and C=Customer.

"Document" Life Cycle Actions	Interoperability & Re-usability Mark-up languages and Schemas	Delivery The internet	Interface Portals	Access & Use Navigation	Management Applications / Actions
Version Control	XML	Intranets		Taxonomies	CMS (D2P)
Retention Schedule	Dublin Core	Extranets		Collaboration	DMS (D2P)
Authentication	DOI				RMS (D2P)
Permissions	OAI				Knowledge Management
Indexing	RDF				e-learning
Classification	Topic Maps				HRM
Thesauri					K-elicitation (P2D)
Taxonomies					K-sharing (P2P)
					e-Commerce
					B2B
					B2C
					CRM

XML = Extensible Markup Language DOI = Digital Object Identifier OAI = Open Archives Initiative RDF = Resource Description Framework CMS = Content management System DMS = Document Management System RMS = Records Management System HRM = Human Resource Management B2B = Business to Business B2C = Business to Customer CRM = Customer Relationship Management D2P = Document to Person P2D = Person to Document P2P = Peer to Peer

Fig. 3. Information Architecture: a road map



area; the former finding it difficult to discover what customers really need, and the latter often blinkered by preconceptions of easy solutions, often over-influenced by what other organisations have done. Information architecture has, as a fundamental principle, the aim to produce the best possible specification for an organisation — one that relates fully and uniquely to its structure, its processes, its other tools, and above all to the people concerned.

The Delphi Group has listed six prerequisites for portal software, currently the prime, and often only, interface between users and information: publishing, application integration, collaboration, personalisation, navigation, and search.

Publishing is a two-edged sword. Clearly, the intranet relies on the ability of authors to easily mount information for access by others, but too many organisations fail to control the process. One multinational has admitted in a private communication to the author that it has 2000 servers connected to its intranet, and that divisional webmasters have lost track of how many webpages fall under their jurisdiction. Publishing standards are necessary to combat this information anarchy, and these may include authorisation processes and indexing by the generators of information. This last consideration has become a key issue in many intranets, where authors are loth to index their own material. Consequently, much thought is being given to ways in which to make it easy for authors to index effectively. One method, being trialled in a U. K. Government Department is to automatically index the document as soon as it has been completed, and to offer possible keywords to the author extracted from a corporate taxonomy. Many organisations are beginning to incorporate templates for authors in a format that mirrors the directory format for searching.

Application integration has been mentioned in relation to figure 3, and is an issue that includes technological and human issues.

Collaboration tools are often offered by the portal software, and may involve a particular use of taxonomies where they have been designed to support, for example, Communities of Interest.

Personalisation is a popular current issue and, as with collaboration, is human-focused. Again, taxonomies may be elaborated to meet specific needs at the individual level, though the implications for construction and maintenance can become problematical.

Finally, search and navigation complete the picture, and whereas these are the core areas of attention for taxonomy building, it can now be seen that they are not the only areas.

One of the reasons why information architecture is so important relates to the huge and bewildering choice of software on the market. Search engine technology has developed rapidly over the last few years, and now includes the following

techniques, all of which are of interest to the information architect, and which may be used in various combinations:

- Boolean
- Probabilistic
- Vector-based
- Linguistic analysis
- Natural Language Processing
- Taxonomies
- Clustering
- Ontologies

Forrester Research (1997-2004) have identified around 100 vendors in the search engine space, who they have classified into three groups:

- Platform players (e. g. IBM, Oracle)
- Industry giants (e. g. Verity)
- Hot upstarts (e. g. Google)

A recent report of research from the Delphi Group (2002a, p. 2), looks at the market share in this area, producing the following figures:

- Basic search: 40%
- Advanced search: 27%
- Taxonomy/Classification: 23%
- Question answering: 8%
- Ontologies/Topic Maps: 2%.

Watching the industry and the way in which it moves both technologically and commercially is an important and time-consuming task for the information architect, but one that is vital if important knowledge is to be available at the specification and procurement stage.

2.4. Processes/Tools

Some of the tools discussed above are for individual information support, while some relate more specifically to people working on shared processes. Such tools can support knowledge sharing by, for example, providing a facility for readers of documents to annotate those documents and pass them to groups or designated colleagues. Other systems will record who has read what, so that peers can solicit opinions regarding the document read. Attempts to relate business processes to information life cycles has also already been mentioned.

2.5. Tools/People

It can never be over-emphasized that tools must be designed for use and that they must be acceptable in all respects, including the physical and the intellectual. This is where Dillon (2000) makes his plea for the importance of an understanding of cognitive psychology in support of the design of human/computer interfaces; and this includes both the visual and usability aspects as well as findability. Personalisation and visualisation are both aspects of this area of consideration. Figure 4 shows a schema for an interface designed by one of the pharmaceutical companies. The master screen at the global level carries a bar along the top, moving from the personal to the organisational. Each drop down menu, powered by a corporate taxonomy, leads the user to information concerning levels of information from

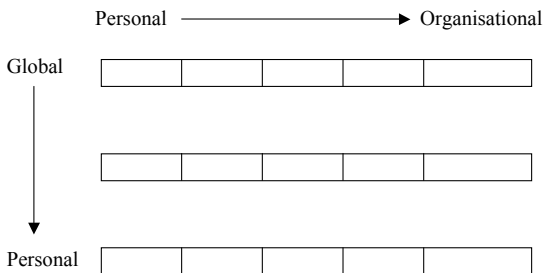


Fig. 4. Schema for a taxonomy map

personal contract details, through opportunities for training and personal development, through work support tools and information, and finally to information about the enterprise itself, mirrored on the Company Internet page. The second bar in figure 4 can be invoked at a national level where American English and concepts have been translated (again through the corporate taxonomy)

into English English. At a third level, the bar can be customised to select only that information which is relevant to a particular site, and finally there are facilities for the individual worker to personalise his access. While such richness will be beyond the means of many organisations, the principles are correct, and may be adapted using other and less expensive resources.

The underlying and basic problem at the junction of Processes and People is language. Humans have the ability to process language at six different levels, an ability that machines can only approximate at the first four, having to receive specific instructions if they are to be able to cope with the other two. The levels are:

- Morphological
- Lexical
- Semantic
- Syntactic
- Pragmatic (e. g. the White House in Washington as a specific white house)
- Discourse (e. g. the use of Picasso, the artist, and he to mean the same person within a passage of text).

Figure 5 shows terminology as the meeting point between the pragmatics of information processing on the one hand: content analysis, particularly subject indexing, and the identification and understanding of user needs; both within the context of the organisation, and a range of sophisticated techniques currently available. The current consensus is that optimal results can be achieved only by combining the pragmatics of human intervention with software support.

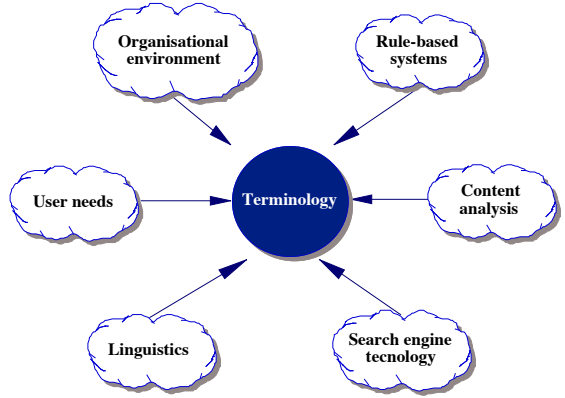


Fig. 5. Terminology as a central issue between pragmatics and advanced software techniques

3. Conclusion

Figure 6 presents a picture of the scope of the ideal corporate taxonomy, bringing together and correlating all the terminologies used by the organisation. In principle such a taxonomy (or taxonomies — all derived from the master taxonomy) could be used to describe all the content of the organisation indicated inside the rectangle by the traditional symbols for documents, but including documents in the widest sense; all the people of the organisation — such description encoded in expertise directories; and finally the organisation itself — its functions and processes, procedures, rules and regulations, standards and forms used. The taxonomy could also cover the related content, people and organisations outside the enterprise.

Such an ideal requires wide-spread co-operation and the ability of the information architecture team to overcome all those human and political problems that

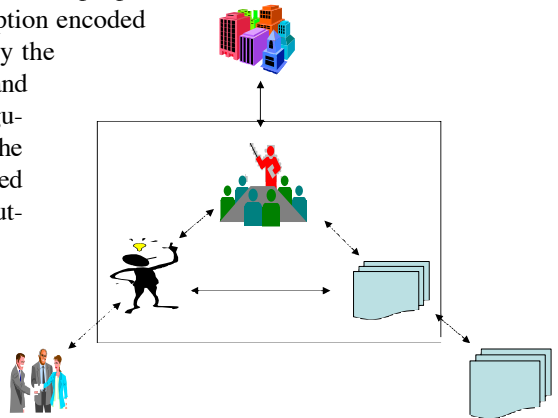


Fig. 6. The total potential scope for a corporate taxonomy

organisations are prone to. The Team must establish principles of governance and ownership. All the stakeholders, including the users should be brought into the picture at the earliest possible opportunity and a Steering Group set up to represent their interests. Above all, and in order to link the proposed implementations to the business, there should be a high level sponsor and champion for the project.

There is much here to engage the traditional skills of librarians, documentalists and information scientists; but only if they become more conversant with the business of the organisation and understand the languages of the different players vying for positions in the multidisciplinary teams of information architecture. They have the potential to assume any position from taxonomists to information architects on a technician/manager continuum.

4. References

- Asilomar Institute for Information Architecture. AIFIA, c2004. URL: <http://www.aifia.org/pg/about_aifia.php>.
- Davenport, Thomas H.; et al. (1992). Information politics. // Sloan Management Review. (Fall 1992) 53-65.
- Delphi Group (2002a). Taxonomy & Content Classification: market milestone report. January 2002. URL: <<http://www.delphigroup.com/coverage/taxonomy.htm>>.
- Delphi Group (2002b). Information World Review. Boston, MA: Delphi Group, November 2002.
- Dillon, Andrew (2000). Information architecture why, what and when. URL: <<http://www.asis.org/Conferences/Summit2000/dillon/index.htm>>.
- Forrester Research (1997-2004). Cambridge, Mass: Forrester Research Inc., 1997-2004. URL: <www.forrester.com>.
- Gilchrist, Alan (2003). Thesauri, taxonomies and ontologies: an etymological note. // Journal of Documentation. 59:1 (2003) 7-18.
- Kelleher, Dominic; Levene, Simon (2001). Knowledge management: a guide to good practice. PAS 2001. London: British Standards Institution, 2001.
- Wiggins, Bob (2000). Effective document management. Aldershot: Gower. 2000.