Terminology Overload?

Terminolojiye Aşırı Yüklenme

Alan Gilchrist*

Abstract

The information industry is now full of professionals from many areas, and those who have been working in that area the longest: that is to say, librarians, documentalists and information scientists, must learn the language that these other professions use if they are to maintain their own professionalism and work effectively (as they must) with the relative newcomers.

The belief in the statement above is exemplified in the new use of the word taxonomy, and several meanings of this word are discussed with reference to findings from a research study conducted by the author.

Keywords: Automatic indexing, Information overload, Taxonomy, Automatic classification.

Öz

Bugün, enformasyon sanayiinde birçok alandan gelen profesyoneller ve bunların arasında bu alanda en çok emek vermiş olan kütüphaneciler, belge ve bilgi bilimciler, eğer kendi profesyonelliklerini sürdürmek ve göreceli olarak bu alana giren yeni kişilerle etkin bir şekilde çalışmak istiyorlarsa, diğer meslekte olanların dilini öğrenmelidirler.

Yukarıda ifade edilen gerçek, "taksonomi" teriminin yeni kullanımına ilişkin bir örnek üzerinde anlatılmış ve yazar tarafından yapılmış olan bir araştırmanın bulgularına dayanarak, bu terimin çeşitli anlamları açıklanmıştır.

Anahtar sözcükler: Mekanik indeksleme, Aşırı bilgi yükleme, Taksonomi, Bilgisayarla sınıflama.

^{*} Cura Consortium and Senior Associate Consultant, (cura@fastnet.co.uk).

Introduction.

The word "taxonomy" has now entered the vocabulary of librarianship and documentation; or rather one should say re-entered, as it is a term well understood by librarians, documentalists and information scientists to mean classification, particularly in the context of Linnaeus. It is therefore clear that the word has been introduced by others. Indeed, this seems to be the case and, furthermore, it is by no means the only word that has been introduced into the sphere of information retrieval, a sphere that librarians and information scientists have, for some time, regarded as their own. There are two major factors behind this appropriation (or rewording) of "our" terminology. The first comes with the emergence of Information Management and more recently Knowledge Management with their emphases on the word 'management'. At last, senior management has woken up to the importance of 'information', and so Chief Information Officers and Chief Knowledge Officers are being appointed in the larger enterprises. However, the former tend to be recruited from the area of Information Technology; and the latter often from Human Relations Development. The second factor is that, for some time now, the vendors of software and information (and to some extent hardware) have targeted the end-user market; a far larger and more lucrative market than that represented by the information intermediaries, and this dialogue has introduced yet more neologisms.

Many librarians and documentalists have long recognized, for example, that enterprises would find it sensible to attempt to relate externally purchased information with internally generated information; or that people are prime repositories of information and that when they leave the enterprise, they take that information with them. Senior management now also recognize these two factors; and, unlike the great majority of librarians and documentalists, have the money and the power to do something about it. And in doing something about it, they tend to turn to the information technologists, the software vendors and the management consultants who all bring their own views – and their own terminologies – to the problem. In many ways, this is a good thing. Some of the old and sometimes obstructive professional barriers are coming down, as for example between librarians, documentalists, records managers and corporate archivists; but there is still a long way to go in bring-

ing these people more closely together to work in more effective partnership with information technologists and senior managers. One of the fundamental problems is that of terminology, of sharing a common language in the workplace. There are some encouraging signs that the skills of librarians and documentalists will not be ignored; so, for example, the well-known portal Yahoo is said to employ some 400 librarians in the task of categorizing the evergrowing numbers of websites (growing at the rate of 7.3 million pages per day!). However, if librarians and documentalists are going to make even more headway, they will have to keep up with a lot of new jargon being coined in the areas of information management, knowledge management and information retrieval. It is not just a simple question of agreeing that an English person should use the word 'sweets' while the American uses 'candies'. This "new" word taxonomy, despite the protestations of some librarians is not the same as classification, though it has strong connections. Nor is it, as some would maintain, just another name for thesaurus; though again there are strong connections. What then are taxonomies in the new sense?

Reasons for taxonomies

The fact that others might introduce old words with new meanings into previously relatively stable domains of discourse should make us think. So, what have been the triggers for this particular development with regard to taxonomies? There are perhaps four obvious factors at work:

- Information overload. Conventional search engines are often now seen to be inadequate in dealing effectively with very large databases, and it is apparent that users need complementary search aids and filters.
- Information literacy. Research has shown that the majority of endusers have severe problems in knowing how to search for information, leading to wasted time and the missing of useful information.
- Organisational terminology. Published classifications and thesauri do not reflect the particular languages of organizations, in which, typically, 80 per cent of the information held has been created internally.

 "Destructuring" of organisations. Mergers and Acquisitions have created big cultural problems at the implementation stage. Similar problems are found in partnering through extranets, and in the establishment and operation of virtual communities.

Types of taxonomies

There are, perhaps, four manifestations of taxonomies in the modern sense (discounting the traditional Linnaean classification). These address different issues, or reasons as discussed above; and may be used in combination.

<u>Web directories</u> : These are commonly used on the Web and are, in fact, a form of classification. A menu of top terms is offered to the user. Clicking on a selected term will display a second level, and so on, for several more levels, arriving finally at perhaps some information or references, or the possibility of offering the last selected term to one or more search engines. Each level does not have to be hierarchical in the normally accepted sense, and terms may be repeated at different levels, providing alternative pathways for the searcher.



Figure - 1

Terminology Overload?

Bilgi Dünyası 2001, 2(1): 37-47

5		Fgworites		Help			-						-
ack	-	- 3	2	6	0		3	12	-				
-	Forned	5100	Planes	h Home pe/United_K	Search	reveren	PEBLORY		PUS .	22	-	€Go Li	1. 7
ions life] hup://an	oz.org/negk	Shat/E URO	pe/United_N	ngoonvso	ciety_and_C	URUFE/FOR	cs/raties/	SOCIANS(_)	aruv		00 10	
acar		-		2.5			-	-	-				
	OLZI OF	en directo	ory proj	ect									
			-		ab	out dmon	add U	RL upd	ate URI	becom	e an editor	feedback	
						-	_				-		
							the	entire dir	ectory				
-													
		Europ	e Un	ited Kin	dom:	Society	and Ci	ulture:	Politic	s: Parti	e: Sodal	lst	
rty (
	1/)							_		_		_	123
	17)		-	- 00				-				_	HILL
		arty (Eng	land a	nd Wales	. T	he party's	man st	e. Camp	augns for	a socialis	t society bas	sed on	THEFT.
• So	ocialist P										t society bas Today". For		THE REAL PROPERTY OF
• <u>So</u>	eting the	needs of	all Pub	lishes wee	kly news	paper Th	e Sociali	st" and n	nonthly "	Socialism		rmerty	TANK COMPANY
• So me kn	ecialist P teting the town as M	needs of filitant La	all Pub bour an	lishes wee	kly newsj tant tend	paper Th	e Sociali	st" and n	nonthly "	Socialism	Today" For	rmerty	THE REPORT OF THE PARTY OF THE
• So me kn ele	ections un	needs of filitant La der the na	all Pub bour an me "So	hshes wee d the "Mill cialist Alte	tant tend mative".	paper "Th ency". Cu	e Sociali arrently h	st" and n as three l	ocal cou	Socialism incillors el	Today". For ected. Stand	rmerty ds in	THURSDAY STREET
• So me kn ele	ections un	needs of filitant La der the na ocialist Pa	all Pub bour an me "So arty - A	hshes wee d the "Mill cialist Alte	tant tend rnative".	paper "Th ency". Cu	e Sociali arrently h	st" and n as three l	ocal cou	Socialism incillors el	Today" For	rmerty ds in	
• So me kn ele • <u>Co</u> Ne	ections un oventry S ellist and	needs of filitant La der the na <u>ocialist Pa</u> Karen Mo	all Pub bour an me "So arty - A :Kay.	hishes wee of the "Mill ocialist Alte voice for	tant tend mative". Coventry	paper "Th ency". Cu 's workin,	ne Sociali arrently h g class, c	st" and n as three l	nonthly " local cou	Socialism incillors el ibour with	Today". For lected. Stand	rmerty ds in Dave	
• So me kn ele • <u>Co</u> Ne • <u>Lo</u>	eting the terms the terms as he terms un oventry S ellist and ondon So	needs of filitant La der the na <u>ocialist Pa</u> Karen Mo	all Pub bour an me "So arty - A :Kay.	hishes wee of the "Mill ocialist Alte voice for	tant tend mative". Coventry	paper "Th ency". Cu 's workin,	ne Sociali arrently h g class, c	st" and n as three l	nonthly " local cou	Socialism incillors el ibour with	Today". For ected. Stand	rmerty ds in Dave	
• So me kn ele • <u>Co</u> Ne • <u>Lo</u> and	orialist P teting the town as A ections un oventry S ellist and J ondon Sou d links.	needs of filitant La der the na <u>ocialist Pa</u> Karen Mo cialist Part	all Pub bour an me "So rty - A Kay. Y - Lor	lishes wee of the "Mili ocialist Alte voice for adon brand	kly newsy tant tend mative". Coventry h of the p	paper "Th ency". Cu 's workin, party. Site	e Sociali errently h g class, c e includes	st" and n as three l opposing downlo	nonthly " ocal cou New La adable p	Socialism incillors el abour with setitions, th	Today". For ected. Stand Councillors he party's pr	rmerty ds in Dave rogramme	
• So me kn ele • Co Ne • Lo ano • Sa	orialist P teting the town as A tetions un oventry S ellist and S ondon So d links. we Free I	needs of filitant La der the na ocialist Pa Karen Mo cialist Part Education	all Pub bour an me "So arty - A :Kay. Ty - Lor Campa	lishes wee d the "Mili ocialist Alte voice for idon brand ign - Cam	kly newsy tant tend mative". Coventry h of the p	paper "Th ency". Cu 's workin, party. Site	e Sociali errently h g class, c e includes	st" and n as three l opposing downlo	nonthly " ocal cou New La adable p	Socialism incillors el abour with setitions, th	Today". For lected. Stand	rmerty ds in Dave rogramme	
• So me kn ele • Co Ne • Lo an • Sa wil	etialist P eeting the own as h ections un oventry S ellist and d links. we Free I th the So	needs of filitant La der the na <u>ocialist Pa</u> Karen Mo <u>cialist Part</u> Education cialist Part	all Pub bour an me "So nty - A Kay. Ty - Lor Campa y (CW.	lishes wee d the "Mili ocialist Alte voice for adon branc lign - Cam I).	kly newsy tant tend mative". Coventry h of the p paign to	paper "Th ency". Cu 's workin, party. Site oppose hu	e Sociali grently h g class, o includes ition fees	st" and n as three l opposing downlo s through	New La adable p	Socialism incillors el abour with estitions, th ic of a ma	Today". For ected. Stand Councillors he party's pr	rmerty ds in Dave rogramme Affiliated	

Figure - 2

Common web directories, of which most, or perhaps all, are manually created include The Open Directory (Figs 1 and 2).

Taxonomies to support automatic indexing : In this case, all that the user sees is a shallow classification of typically two levels and perhaps 1000 to 4000 terms in all. What the user does not see is the rules base behind the classification, which contains sets of terms, weights and instructions relating to each term; the rules base being used to automatically extract appropriate index terms, which may, or may not, be present in the documents. This approach is particularly attractive where throughputs are so large that manual indexing is economically impossible. Nevertheless, the compilation of these rules bases is very labour intensive, often taking 4 or 5 hours per term.

BBC Rules Base

Marithon (gypsies and travellers BBC off) File Pri Antes New Indian Ring		
		and the second s
# O 100 BBC.GYPSIES-lates (Accure)[0]		
III-O 100: BBC.GYPSIES-Mer.2 (Accrue)[0]		
B 100 BBC GYPSIES-artefacts (Accrue)[0]		
5 cMany/ (Stem) caraven [3]		
O 10: (Maryo (Phrase) [g7ps' wagons] [0, 0]		
# O 10: (Many) (Phrase) [home diawn caravana] [0, 0]		
O 10. (Manip (Phrase) (horse drawn vardos) (0, 0)		
5 (Margo Bananco (0)		
5 (Mary) varios (0)		
# 100 BBC.GYPSIES-culture (Accrue)[0]		
B 100 BBC GYPSIES-culture2 (Accrue) [0]		
at O 100 .BBC.GYPSIES-culture-gypties (Accrue)[0]		
9 00. BBC.GYPSIES-history (Accrue>[0]		
a 100 BBC GYPSIES incidents (Accrue)[0]		
100. 88C GYPSIES-orderv groutes (Accrue)(0)		
100 BBC GYPSIES onone general (Accrue)(0)		
a 0 100 BBC GYPSIES origina roma (Accrue>101		
O 100 BBC GYPSIES man mna2 (Accrue)[0]		
# • 100 BBC GYPSIES signe traveless (Acorum (0)		
# O 100 BBC GYPSIES-people (Accrue)[0]		
IDD: BBC GYPSIES-symotymes (Accrue)(0)		
# • 100 BBC GYPSIES traveliers (Accrue)[0]		
B- 100: BBC, GYPSIES-wander (Accrue)[0]		
and the data of the second states of the	U.S.	and the second s
1045 Straw I won't back down on gepties	/182 082	C \My Documents\T oby Million/BBC\DumpFiles\tavefiles and
1061 No room at the site for ministrad baveling people 1049 Anger at Staw's Gypsies remark	089	C:\My Documents\Toby Milner\88C\DumpFiles\havelless.co C:\My Documents\Toby Milner\88C\DumpFiles\bavelless.co
1043 Angeral Staw's bypeen remain. 1006 Straw's travellers galle 'minconstrued'	085	C \My Documents\T oby Miner\BBC\DumpFies\travelers.x
2006 Straw's traveliers gate inisconstrued	085	C:\My Documents\Toby Miner\BBC\DumpFiles\travefers.xx
050 Meking sense of the rules of the open road; Law	095	C \My Documents\Toby Milner\BBC\DumpFiles\travellers.xo
0058 Invasion is leared as Blair opens country to gapties	085	C VMy Documents/Toby Milner/88C/DumpFiles/Itavellers.to
**	11	Harts I
		http://www.itic.com
Statt DE Enterry - C'My Docarra		1

Figure - 3

Figure 3 shows part of a rules base employed by the British Broadcasting Corporation in its NEON (NEws ONline) information retrieval system. This was built by knowledge engineers in the software house Smartlogik, using Infosort programs.

<u>Taxonomies created by automatic categorization</u>: A growing number of software packages are now available which purport to be able to analyse text, to automatically create categories from that analysis and to classify the analysed documents according to the categories created. These categories may then be displayed as web directories (see above) and/or as two- dimensional maps, where related terms are linked to the selected term that appears

in the middle of the map. Selection of a related term will then move that term to the centre and introduce a new set of related terms. In practice, these packages are effective only when dealing with relatively large collections of conceptually homogeneous material, and when seeded at the beginning of the operation by loading relevant terminologies or glossaries. They also normally require a significant amount of human editing and maintenance.





Figure 4 shows a web directory menu automatically generated by software supplier Semio from a body of text within the United States Postal Services, internally seeded by loading external and internal glossaries of business terms.

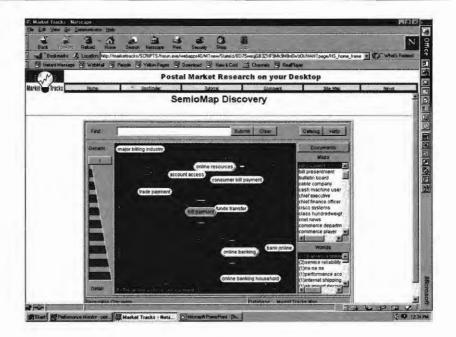




Figure 5 shows the alternative display of a map.

The preceding three examples of different taxonomies are addressed either at the processing stage of information input, or at the display of results. However, the modern enterprise faces the whole range of problems listed above and must consider all the new approaches being promoted by the suppliers. The prime need to make information easily accessible to its staff through the Enterprise Information Portal or other channel is often compounded by large throughputs and legacy data. Increasingly, there is also pressure to provide maps and user guidelines to the repositories and their contents, which may be numerous and include both internal and external sources. This may introduce the deceptively simple-sounding word 'mapping' that, in fact, indicates the need for a huge amount of intellectual human effort. The need for considerable human intervention was a major finding from the research conducted last year by TFPL Ltd., as was the multiplicity of information architectures employed by the large organisations [1].

<u>Corporate taxonomies</u> : One manifestation of such different architectures includes a super-thesaurus, as used by Glaxo Wellcome (now GlaxoSmithK-line) wherein a number of existing thesauri have been loosely merged and each term carries the addresses of any corporate information repositories using that term (Figures 6 and 7).

Therachical display of 'INTERFERON' Microsoft Internet Explorer	一百日日
A R R A R R R R R R R R R R R R R R R R	
Dure: april 2017 Derweitig Theseast Kada 1901 Pacifics Käning Off	
Tops: © CHEMICALS AND DRUGS (MA(top)) © CHEMICALS, DRUGS AND MATERIALS (M(top))	000
BTs: Cricolane (E.S. (H) Anterioral Assent (E.S. (H) E.S. (H) Mattheway (E.S. (H) Matthe	
Term:	6
E State Interferon (EA) cl SS4 (EA) endogenous statefaron (EA) exogenics interfaron (EA) in (EA)	
un (EA) interferon 1 [EA]	
)	I Munimume .
Hent architela dantag a	10 Th 1.







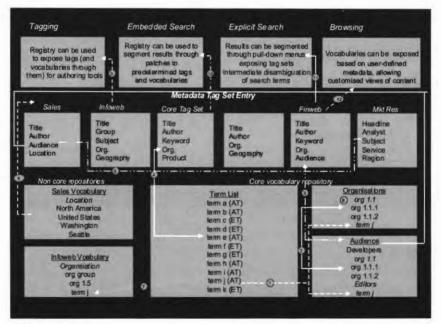


Figure - 8

This is a huge effort with some 53,500 basic concepts (or lead terms), 201,750 synonyms (remember, this is the pharmaceutical industry with trade name and chemical equivalents round the world), and 443,500 related terms; another is a corporate macrothesaurus containing terms common to the whole enterprise, and to which are hooked microthesauri for specialist areas, an approach used by one of the world leader management consultants; a third is a metadata registry made available by the Microsoft Knowledge Architect Team to portal owners and information providers, containing tagging standards supported by generic terminologies (Figure 8 displays the schema for this system); yet another approach provides high-level "knowledge maps" pointing to the location of relevant repositories, with guidelines on their contents and access paths and protocols; finally there are expertise databases supported by the facility to annotate shared documents or a process of collaborative filtering.

Conclusion

This short essay has attempted to show that librarians and documentalists are well equipped to understand many of the new words being introduced into their traditional area of work by other professions; but it is vital to understand the nuances, and the ways in which the new terminologies reflect the dominance of information technology and the focus on business processes brought about by the concentration on knowledge management and the desire to exploit intellectual capital. The single word 'taxonomy' was used as an example, but there are many more words to be assimilated, and doubtless, there will be many more to come.

References

 Gilchrist, A. ve K. Peter. (2000). Taxonomies for business: access and connectivity in a wired world. London. TFPL.