



A TERMINOLOGY-BASED RE-DEFINITION OF GREY LITERATURE

***Claudia Marzi
Gabriella Pardelli, Manuela Sassi***

**Institute for Computational Linguistics (ILC)
National Research Council (CNR) - Italy**



GREY LITERATURE DEFINITION

- The Luxembourg Convention on Grey Literature held in 1997 offered the following definition of Grey Literature (expanded in New York, 2004):

“Information produced *and distributed* on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing, *i.e. where publishing is not the primary activity of the producing body*”.



OTHER DEFINITIONS (I)

- ❑ GL had been for many years synonymous with *reports literature* : from the beginning of 20th century *documents evolving out form research and development, particularly from the aircraft and aeronautics industries were a very important means of communicating the results of research testing* (C.P. Augur 1989)
- ❑ In the '70s GL was recognized as *dissemination vehicle for many organizations [..]* though not easy to find (C.P. Augur 1989)
- ❑ In the '90s GL became a necessary component of effective literature searching and *attained its importance as a separate medium of communication because of an initial need for security of confidentiality classifications which prevent documents being published in conventional manner* (C. P. Augur 1989)



OTHER DEFINITIONS (II)

- ❑ P. Hirtle (1991) defines GL as *the quasi-printed reports, unpublished but circulated papers, unpublished proceedings of conferences, printed programs from conferences, and the other non-unique material which seems to constitute the bulk of our modern manuscript collections.*
- ❑ IGLWG (1995) defines GL as *"open source material that usually is available through specialized channels and may not enter normal channels or systems of publication, distribution [..]"*
- ❑ M. C. Debachere (1995) describes GL as *a range of materials that cannot be found easily through conventionally channels [..] but which is frequently original and usually recent.*
- ❑ Wikipedia (2010) quotes *"Grey Literature is a term used [..] to refer to a body of materials that cannot be found easily through conventional channels such as publishers [..]"*

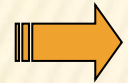


RESEARCH QUESTIONS

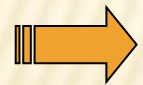
- ❑ Is the Luxembourg definition still valuable?
- ❑ Is it completely satisfactory?
- ❑ Considering other conventionally accepted definitions, does the idea of GL as information produced and distributed by non-commercial publishing take into consideration the increasingly availability and the growing importance of computer-based encoding as the standard medium of creating and developing GL?
- ❑ ...or do the definitions rather need modifications?



OUR GENERAL IDEA



An interesting re-definition of GL can be based upon careful examination of the longitudinal trend of more than 20 years of terminological creativity in the 11 GL International Conference proceedings.



A domain-specific document repository offers the possibility to pave the way to a bottom-up redefinition of Grey Literature stemming from attested terminological creativity and lexical innovation.



REFERENCE CORPUS

- The Corpus of *GreyText Inhouse Archive*, available on <http://www.greynet.org/opensiglerepository.html> represented our empirical basis, consisting of:

titles

themes

keywords

full abstracts

for a total amount of almost seventy thousand word tokens.



METHODOLOGY

- ❑ We focussed on a set of automatically-acquired terms (both single-word and multi-word terms) obtained by subjecting our reference Corpus to a number of pre-processing steps of automated text analysis, such as concordances, frequency lists (selected from around 70,000 words with a higher frequency than 9) and lexical association scores (e.g. *Mutual information* on word pairs).
- ❑ Through manual inspection/classification of relevant list of concordances, we analyse the most relevant terms.



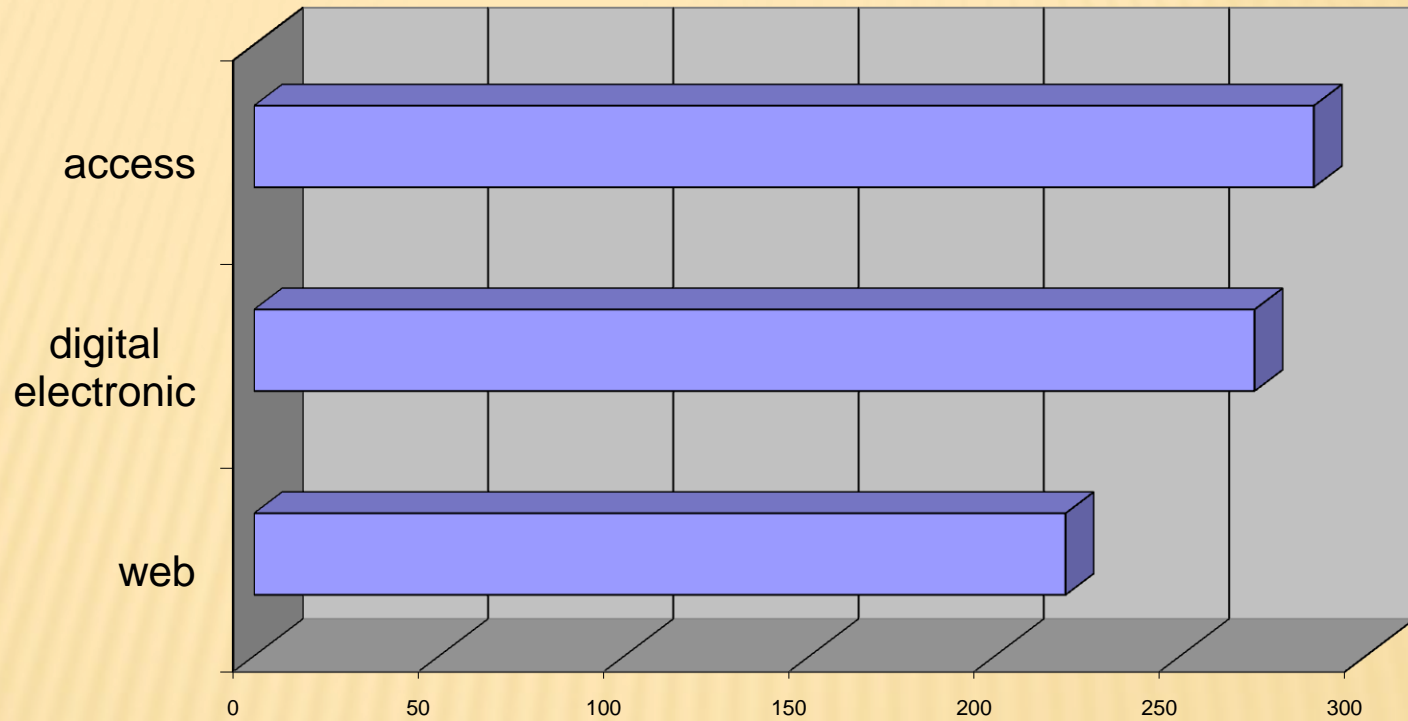
DATA EXTRACTION

Word frequency list

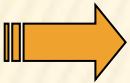
| | | | |
|--------------------------------|--------------------------------|---|----------------------------|
| 985 grey | 126 report | 88 published | 68 health, projects, world |
| 966 literature | 124 available | 85 database, publications, web | 67 databases |
| 737 information | 121 libraries, open, technical | 84 academic, document | 65 community |
| 477 research | 120 data, national collection | 83 analysis | 64 bibliographic, theses |
| 220 access , conference | 106 public | 82 authors | 63 survey, systems |
| 204 library | 104 knowledge | 80 communication | 61 european, sigle, social |
| 202 paper | 103 publishing, university | 77 management | 57 content |
| 191 documents | 102 study | 75 services | 40 accessible |
| 182 scientific | 100 results | 74 countries | 32 online |
| 178 proceedings | 98 international | 73 researchers | 30 www |
| 162 development | 97 resources, technology | 72 government, internet , repositories | 26 accessibility |
| 159 project | 95 metadata | 70 work | ... |
| 142 electronic | 92 repository | 69 materials | |
| 137 system | 91 users | | |
| 130 use | | | |
| 128 science | | | |
| 128 digital | | | |



RESULTS (I)



RESULTS (II)

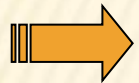


We then considered lexical association scores between salient terms, focussing on terminological usages that are closely related to the mentioned ontological entities.

- ❑ **access:** *easy, electronic, facilitate, full, grey, information, internet, journals, literature, materials, movement, multicultural, open, public, repository, research, scientific.*
- ❑ **digital:** *document, grey, library, literature, network, object, project, repository, system, technology, theses.*
- ❑ **electronic:** *format, grey, information, journal, literature, network, paper, publication, report, resource, technical, theses.*
- ❑ **web:** *access, database, grey, information, network, literature, science.*
- ❑ **knowledge:** *base, exchange, generation, infrastructure, management, scientific, service, share, society, and, especially, information.*



RESULTS (III)

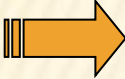


Digital/electronic, access, web mark in our view important stages in the evolution of our current understanding of GL :

- ❑ the attribute *digital/electronic* emphasises the growing importance of computer-based encoding as the *standard medium* of GL;
- ❑ the noun *access* (defining the process of accessing text documents) together with adjectives like *easy, full, grey* and *open* shapes up important conceptual innovations in the way GL material is distributed (*open access* focuses on the free accessibility of digital contents); coupled with *information, document* and *repository*, it points to a conception of world-wide available, structured cultural contents;
- ❑ reference to the *web* lays emphasis on the huge importance of the www as the standard means of disseminating GL.

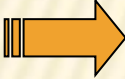



CONCLUDING REMARKS

 GL defines an innovative approach/methodology for information dissemination and exchange

- ❑ web-based sharing and distributed access to openly available scientific/technical document repositories

key notions

- 
- ❑ medium: electronic/digital files
 - ❑ distribution channel: internet/web
 - ❑ information access policy: distributed, unrestricted
 - ❑ access tools: repositories, web browsers, search engines

 Promotion of new, creative modes of production and use of knowledge!

