



A PUBLIC OPEN-SOURCE ENVIRONMENT FOR A SAFER INTERNET ACCESS

WWW.POESIA-FILTER.ORG



Information Society



Overview of Poesia Project

Mohamed Daoudi
Enic Telecom Lille 1

Final Workshop
Pise 21-22 January 2004



- Istituto di Linguistica Computazionale (Italy),
- CEA (France)
- Fundació Catalana per a la Recerca (Spain),
- Enic Telecom 1 (France),
- Universidad Europea de Madrid (Spain),
- Sheffield University (UK),
- Liverpool Hope University College (UK),
- M.E.T.A. S.r.l. (Italy),
- Pixels (Italy)
- Telefónica Investigación y Desarrollo S.A. Unipersonal (Spain).



General Overview

- **Aims of IAP2117 project**
- **Open-source project**
- **Multi-disciplinary filtering project**
- **Technologically Innovative, so requires R&D effort**
- ***Mostly Content* filtering**



Open-source (1)

- Open-source = free (libre) software
- 0. Freedom to **run** the program
- 1. Freedom to **study** how the program works and adapt to your needs (requires access to source code)
- 2. Freedom to **redistribute** copies
- 3. Freedom to **improve** the (source) program & publish improvements



Open-source (2)

- *Poesia* **re-uses** lots of existing OSS (firewall, proxy, MTA, libraries, database...)
- *Poesia* **produces** re-usable software bricks (outside filtering goals) :
 - Image processing
 - NLP processing
 - Generic libraries
 - etc....



Open-source (3)

- Most partners was **new** to OSS
- The **open-source** character of Poesia is one of its **main attractive** features
- Poesia **development** follows **established OSS approach**: CVS (= concurrent versioning system), **mailing lists, Web, ...**



Open-source (4)

- Open-source software has a very **specific dissemination process:**
- Dissemination is (in part) **self-organizing & emerging**
- OSS community believes only in *software that works*



Open-source (5)

- This IAP2117 project is building **tools for filtering**.
- **Integration** of these tools for specific filtering solutions suited to local needs is a **significant task**
- New service **market opportunity** for SME (sys. & network admins...)



Open-source (6)

- Poesia can run on an entirely **free** software system: Linux PC
- Could be integrated into specific filtering boxes (for classrooms, ...)
- Trends **common** to other OSS: Apache (web server), MySQL (RDBMS), etc...



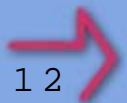
Multi-disciplinary

- **Poesia mixes several technologies & expertises:**
 - Users needs analysis
 - Image analysis
 - Natural language processing
 - Network & system
 - URL, PICS, Javascript



Multi-disciplinary(2)

- Internet **filtering** is an extremely **difficult** & ill-defined subject
- **Fully effective** automatic filtering is **technologically inachievable** (even with much larger means & delays – e.g. NSA)
- Defining **realistic** detailed specifications is already difficult (needs users ↔ developers cooperation)





Multi-disciplinary(3)

- Poesia is mostly a **content** filter, so requires automatic analysis of content;
- But Poesia is also a URL based filter;
- Huge importance of **experimental** development & adaptative techniques



Muti-disciplinary(4)

- Poesia provides a **filtering** system **extensible framework** : extensions to new channels and/or domains should be easily done, even by third parties
- Poesia project should initiate a momentum to **enhance** the Poesia system **later** on.