Towards a General Enterprise Information Logistics Framework

The EILF Proposal
The Current Situation

• Unstructured information management becoming more and more important
• Different vendors, different platforms, different technologies, different methodologies
• No common standard framework available to cover all aspects from data sources to semantic annotations

➡ Proprietary solutions
  ♦ Hard to implement, to maintain, to extend
  ♦ Re-inventing the wheel all the time for commodity functions
Mission

• Establish the future standard and technological infrastructure enabling the next generation of information management systems
• Define and implement a standard framework in the area of (unstructured) information management on enterprise levels
• Establish a community to bring in related aspects, such as
  • OCR
  • data mining
  • language translation
  • rich media management
Strategic Goals

• Develop standard framework for unstructured information management
• Provide ready-to-use platform and components
• Allow for extension in terms of specialized technologies, components and plug-ins
• Establish lively community to get best-of-breed tools in
• Open source to allow skilled developers to contribute
• Allow for business being generated on top of EILF, e.g. via dedicated components or services
Architecture

Key Ideas

- Crawlers/Agents push data into Connectivity / Entry Point
- Connectivity Module filters, converts versions, extracts binaries etc. and pushes into queue
- Message-driven queue stores data and guarantees delivery
- 1...n servers respond to messages, process data and write back to queue
- Potentially multiple instances of servers for load balancing and increased throughput
  - Open issues: synchronization of persistence
  - 1...n processes inside server arranged via BPEL (=pipelines, ~strategies)
- Search yet to be defined separately but the objective is to separate the processes of
  (a) filling the index and (b) using the index for search (unlike in eLIS)
Scope

- Architecture
- Componentization
- Data Source Management
- Operations and Management
- Authentication and Authorization
- Application & Business Use Cases
Team and Community

• Joint initiative by empolis and brox (Germany)
• Key contacts:
  ❖ August Georg Schmidt (brox)
  ❖ Igor Novakovic (empolis)
  ❖ Mario Lenz (empolis)
• Currently ~8 active developers involved
• Resources
  ❖ http://www.eclipse.org/proposals/eilf/
  ❖ Newsgroup eclipse.technology.eilf