

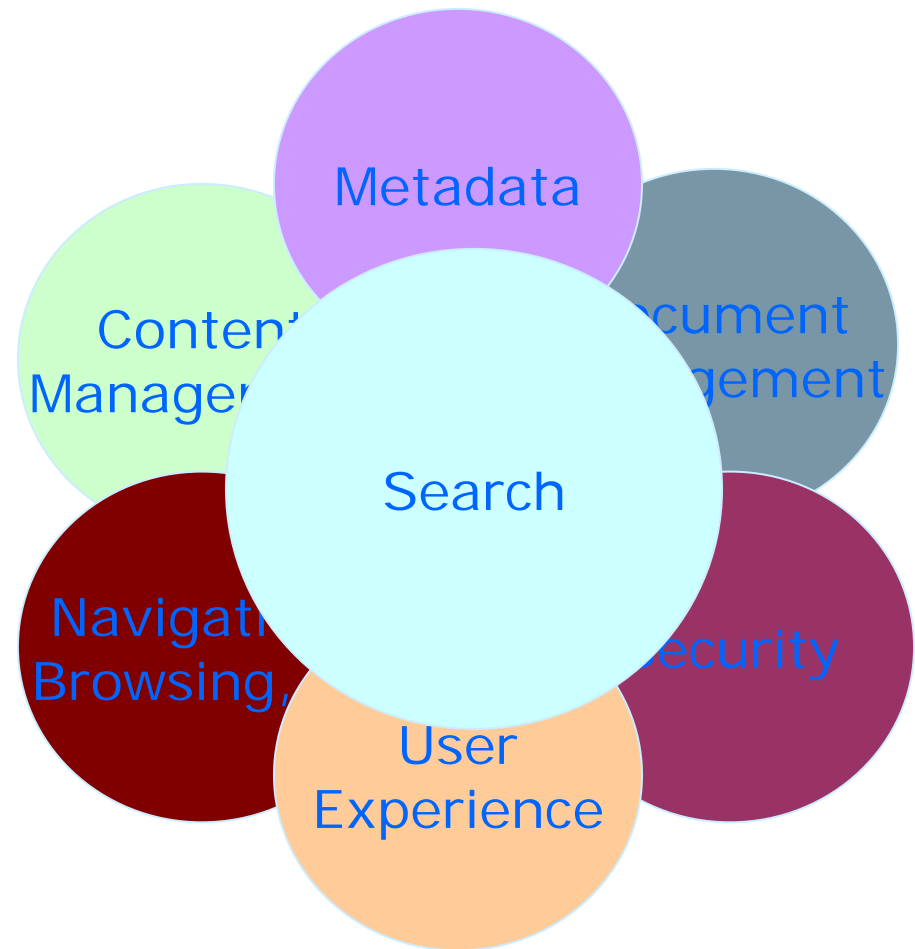
Volocom approach to Enterprise Search

- Enterprise Search architecture
- User needs
- Project approach

- **Users can't find what they need**
 - “I didn't know your department was developing a product similar to ours”
 - “We spend lot of time to find references and case studies to use in pre-sales”
 - “Our sales and support staff give our customers inconsistent information”

Enterprise search

- **Single comprehensive search system implemented within a business context for internal users**
 - Index internal content
 - Serve internal clients and end-users
- **Why is different?**

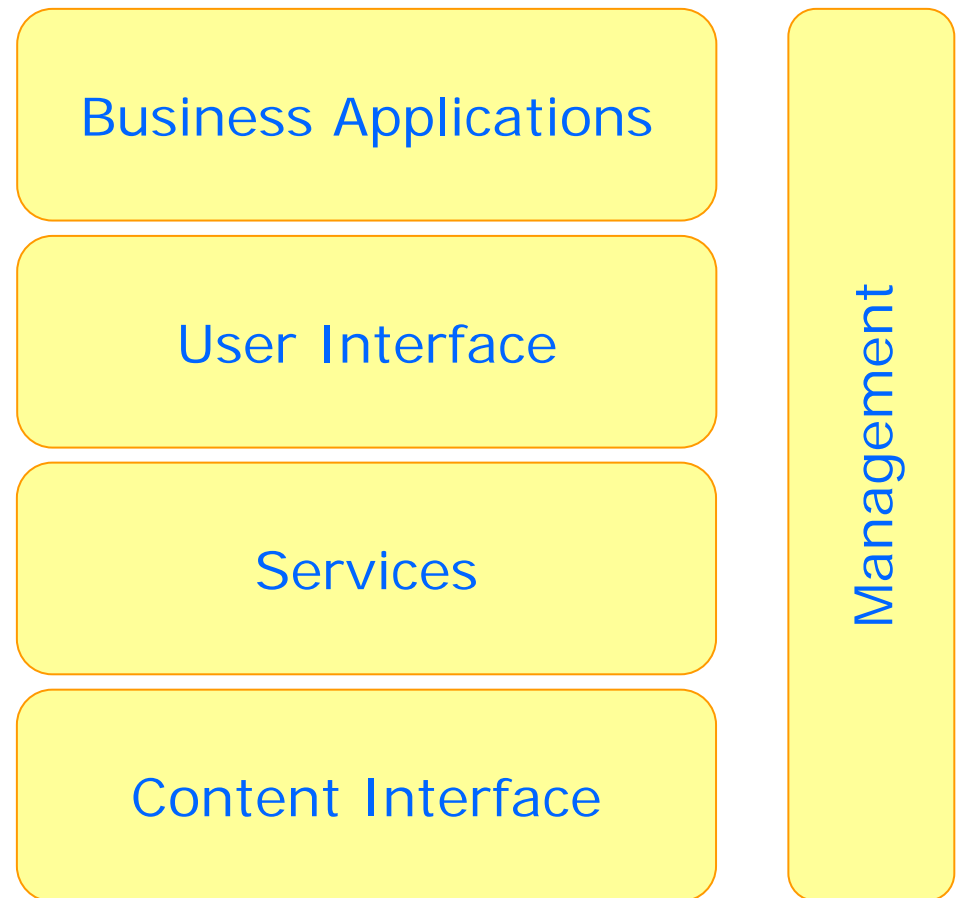


Why it's different?

- **Search is not the business of the enterprise**
- **User task is different**
 - Internet: find an answer or a starting point
 - Intranet: find the right answer, know it's the right answer
- **Ranking criteria**
 - Statistical data may not be significant
 - In-links are not there

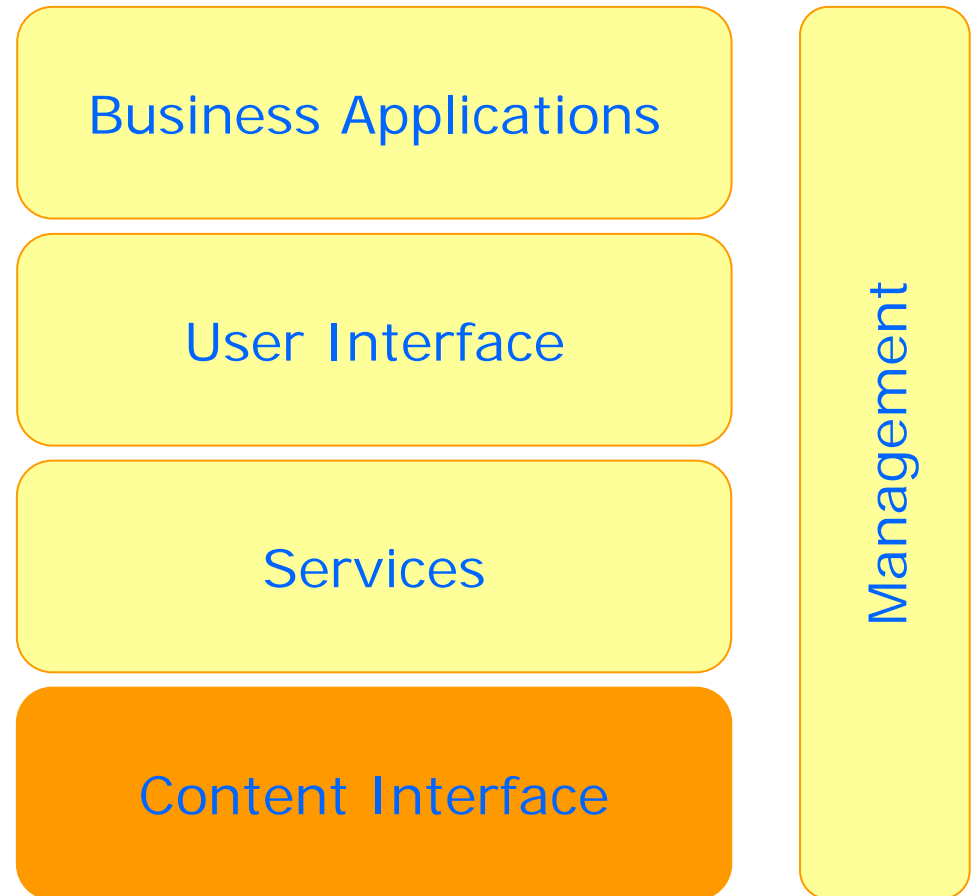
E.S. Architecture

- SOA compliant
- Easy configuration
- High scalability

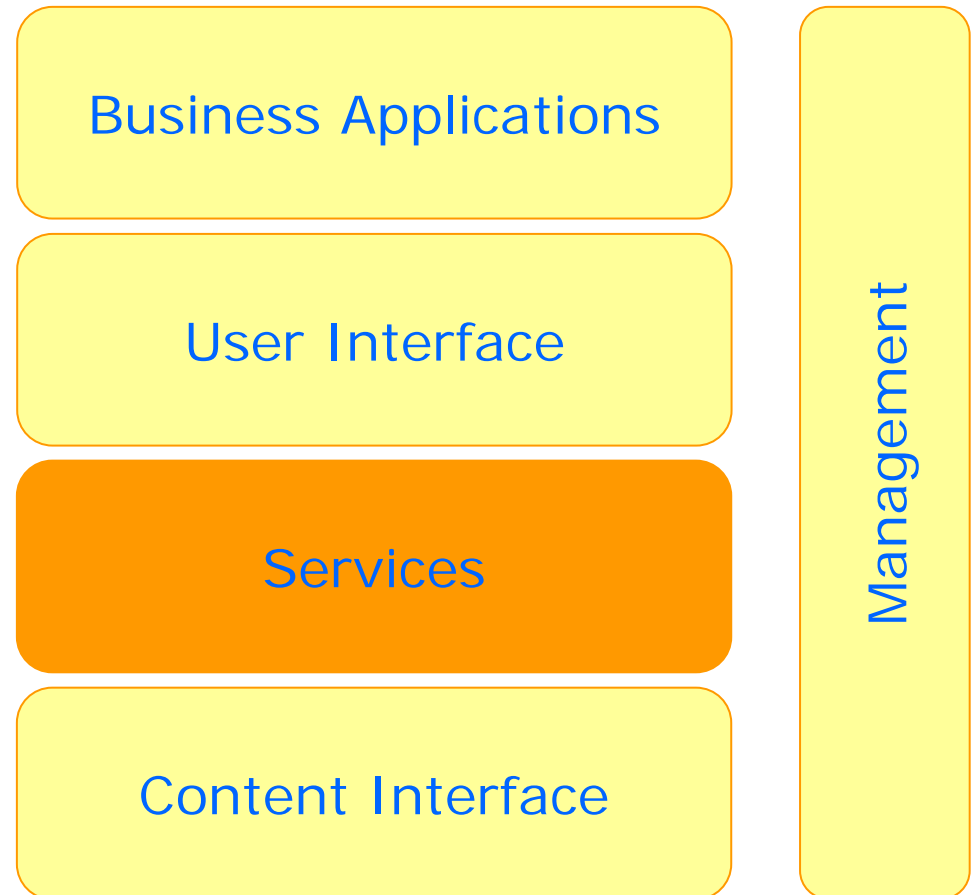


Content Interface

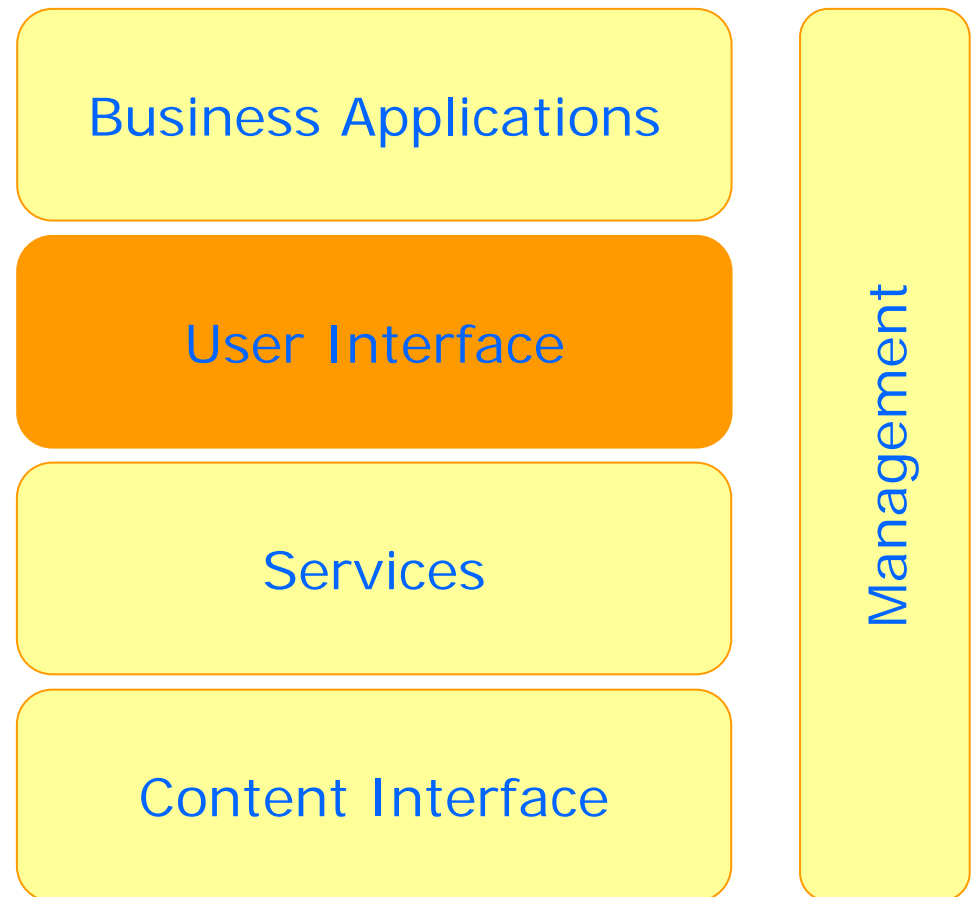
- **Content types**
 - Structured or unstructured data
 - Multimedia
 - People
 - Geospatial, etc.
- **Content Sources**
 - Web, ECM, Apps, Email, Collab, etc.
- **Security**
 - Authorization, Authentication, SSO, etc.
- **Indexing**
 - Federation, Geographic distribution, replication, etc.



- **User-defined classification**
 - Metadata
 - Best Bet
 - etc.
- **Auto-generated classification**
 - Metadata extraction
 - Auto-summarization
 - Entity and concept extraction
 - etc.
- **Clustering**
- **Taxonomies**



- **User Experience**
 - Query UI
 - Keywords
 - NLP
 - Exact matching
 - Results set UI
 - Result set tuning
 - Ranking (relevance, social ranking,...)
- **User Profiling**
- **User Context**
 - Role, Group, etc.



Business applications

- Services that call or embed search
- Industry applications that incorporate search
- Applications on top of search
 - Analytics
 - Knowledge Discovery
 - Workflow
 - etc.

Business Application

User Interface

Services

Content Interface

Administration & reporting

- **Reporting**
 - Standard reports
 - Custom reports
 - Report builders
- **Analytics**
 - Click-through URLs
 - Top keywords, sites
 - Trend analysis
- **Management tools**
 - Scripts and automation
 - Site discovery
 - etc.

Business Applications

User Interface

Services

Content Interface

Management

User Needs

"Moore's Law"

An information retrieval system will tend not to be used whenever it is more painful and troublesome for a customer to have information than for him not to have it.

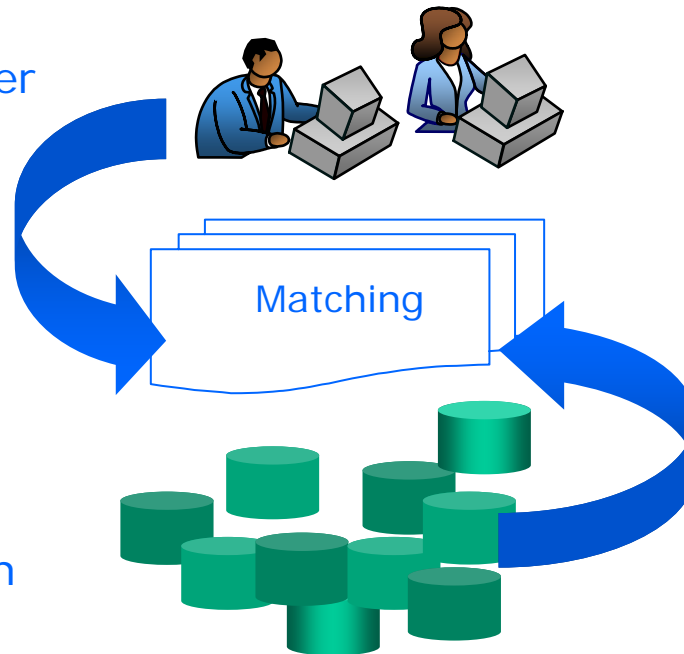
Information need

■ Request

- The terms a searcher uses are not the same as the words the author used
- Term ambiguity
- Misspellings
- Use of acronyms
- Jargon

■ Context

- Query interpretation
- Query expansion



Results presentation

■ Documents ranking

- By relevance
- Social ranking
- Custom

■ Results from different content sources

■ Search refining

- Narrowing
- Expanding
- Sorting

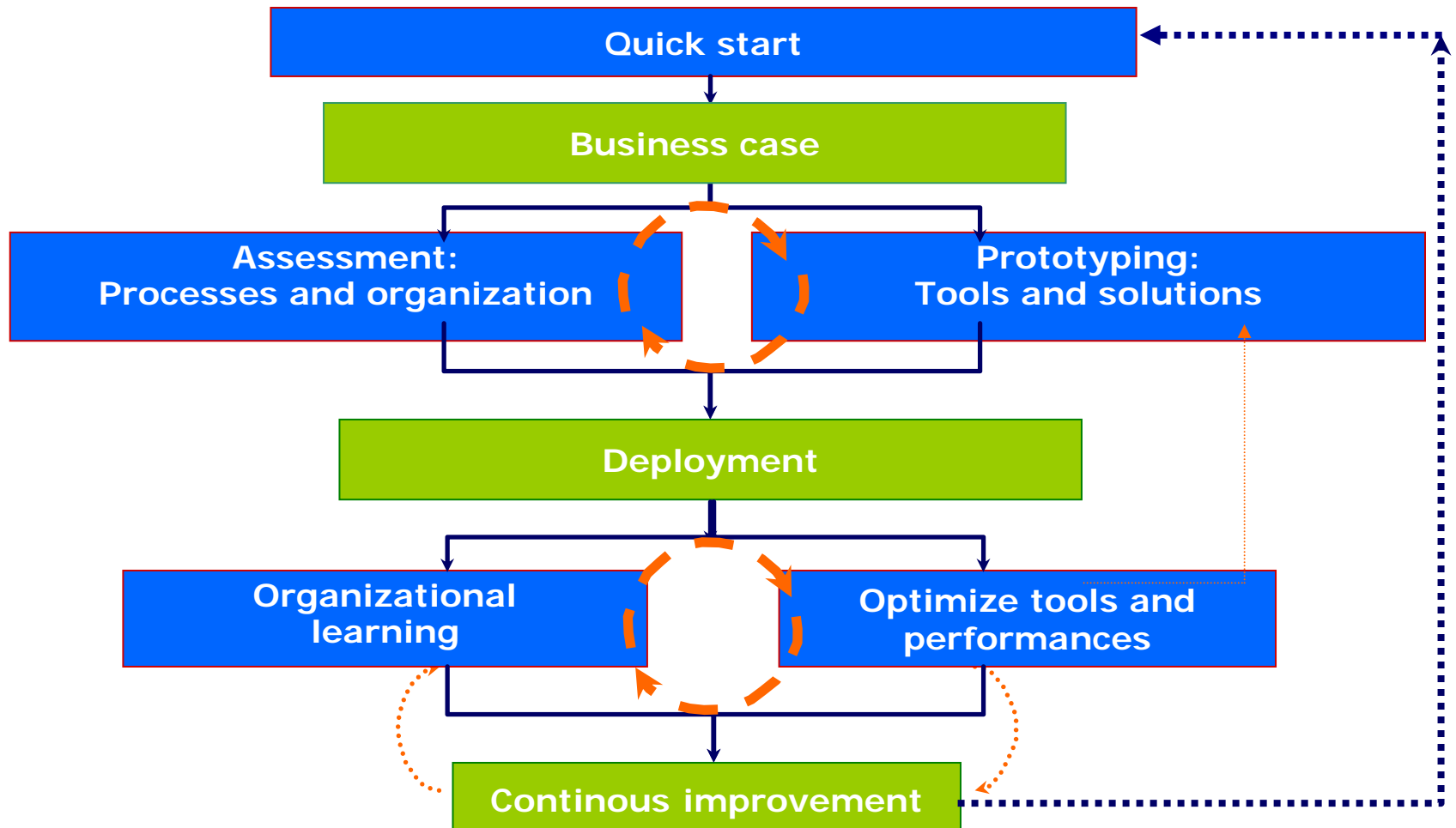
- Can users find the search box?
- Are we indexing the right information?
- Are we protecting information that should not be searchable?
- What can we do to improve the query?
- What search terms are people using
 - that are not returning results?
 - that are returning too many or inappropriate results?
- What next steps can we offer the user if their initial results do not answer their question?
- Which services are we able to offer in order to organize results (e.g. classify them)?

Biggest challenge: Content

- Disparate repositories
- Different technologies
- Distributed systems
- Different releases
- Managed by different groups
- Different navigation and metadata schemes
- No publishing policies

- Different scenarios
 - The content shouldn't be in the search indexes at all
 - The content should be in the index, but only be returned for certain users
 - The content should be returned in results for everyone, but gaining access to the full document should be restricted
 - The content should be returned and accessible to everyone

Project approach



Project guidelines

- **Organizational information strategy**
 - Define business cases
- **User centric approach**
 - Understand user expectations
 - Leverage search to make organizational complexity transparent to the users
- **Audit content**
 - Quality
 - Workflow
- **Enabling technologies**
 - Price/Performance
- **Operational Best Practices**
 - Information architecture
 - Search logs

Return on Investment

- ROI is always a healthy question and it is always difficult to answer
- Faced to exponential growth of data and content, enterprise search is just part of corporate information “infrastructure”
- “Search” is one core activity for “Knowledge Workers” and knowledge workers productivity is a strategic issue for all organization

IDC estimates that an enterprise employing 1,000 knowledge workers wastes at least \$2.5 to \$3.5 million per year searching for non existent information, failing to find existing information, or recreating information that can't be found. The opportunity cost to the enterprise is even greater, with potential additional revenue exceeding \$15 million annually.