WebSphere software

IBM

e business software

Search

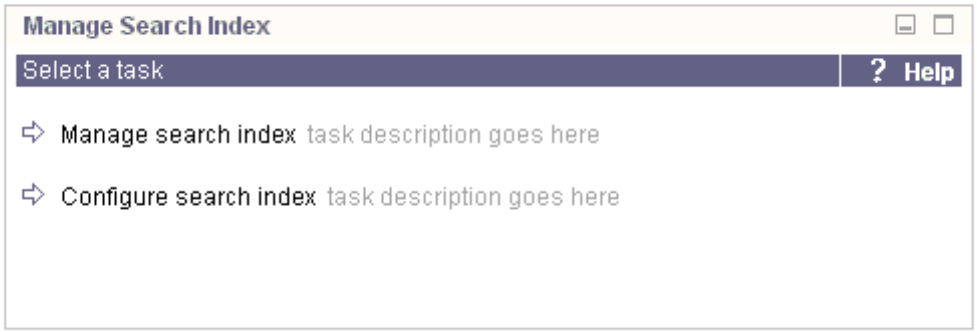
WebSphere Portal 4.1

IBM Software Group

Integration into WebSphere Portal

- **New in WebSphere Portal v4.1.**
- **Embedded search engine available in all offerings.**
- **Search packaged as a portlet application.**
- **Index files on same machines and other servers.**
- **Component portlets:**
  - Administration
    - Build/maintain indexes of content
    - Documents retrieved using a crawler
    - Details on following foils
  - Search and result list display
    - User search requests displayed and ranked
    - View document by clicking on link - displayed in separate browser window
    - Details on following foils

Entry admin portlet:



WebSphere software

New Search Technology

- **Search technology from IBM Research Lab.**
- **Full text search engine written in pure Java.**
- **Allows for sophisticated creation of index.**
- **Language indexing:**
  - Standard indexing for Latin 1 type languages (i.e. SBCS)
    - Languages allowing for word-based tokenization
    - Tokenization rules are based on known word separators.
  - DBCS - uses n-gram indexing
    - N-grams are sequences of n consecutive characters in a document.
    - N-grams are generated from a document by sliding a "window" across the document's text, moving it one character at a time.
- **Used by IBM in "Help Now!" application and others.**
- **Performance information shows that the search technology attains search performance comparable to well-known search engines.**

Components of Text Search

- **Indexer**
  - transforms given documents into a form suitable for allowing fast searches
  - primary keys are "words" and not URL
  - statistics on words kept for relevance scoring
  - abstracts the document
- **Search**
  - UI is a portlet
  - requires indexer as prerequisite
  - types of searches:
    - simple search: uses a keyword or list of keywords and matches those against a document collection
    - advanced search: focus the search on a subset of documents
      - i.e. "must" or "must not" appear in every document

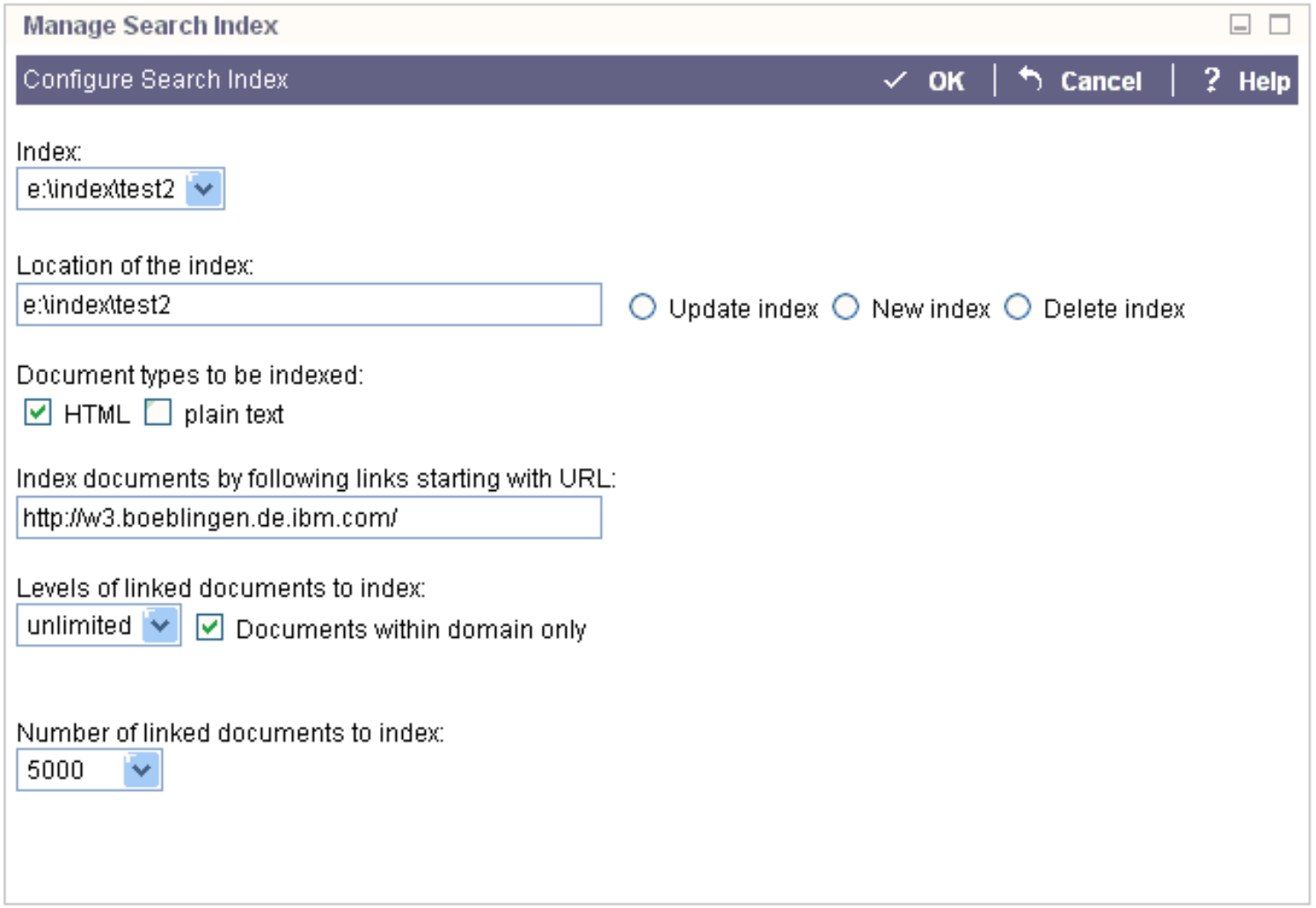
Administration - Index Configuration

- **Administration Portlet**
- **Administrator provides as input:**
  - Unique index identifier or name.
  - Where to store the index.
    - Multiple indexes cannot share same location.
  - Language for optimization
    - Decrease the size of the index by applying stopword processing.
    - For DBCS languages, the appropriate tokenization algorithms can be applied for those languages.
  - Types of documents to index
    - Adobe PDF format not supported.
  - Starting URL
  - Nesting level
    - It is assumed important documents are located at a higher level.
  - Restrict to domain
  - Maximum indexing size
    - Index not yet built until Index Administration performed

Administration Portlet

- **Sample search index configuration**

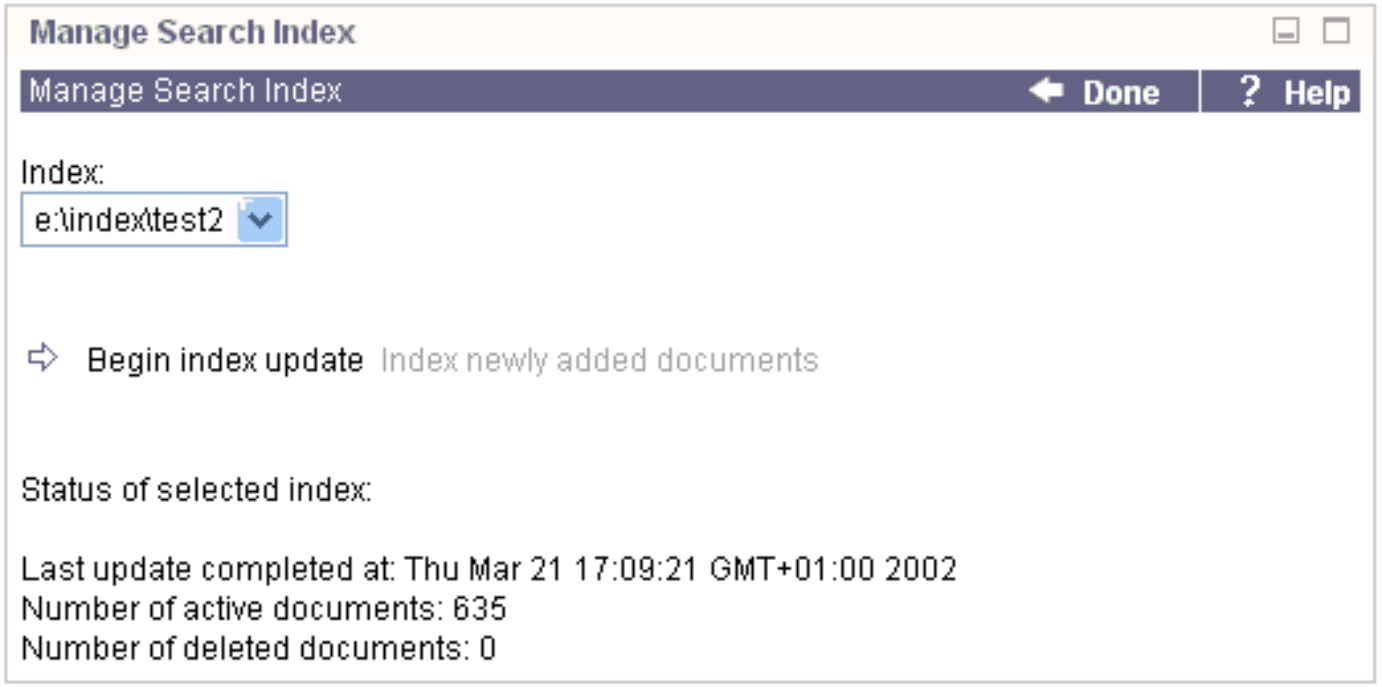
Configure search index



Administration - Index Management

- **Administrator Portlet:**
  - Point and click interface.
  - Points to a configured index.
  - Builds the index.
  - Compaction occurs automatically

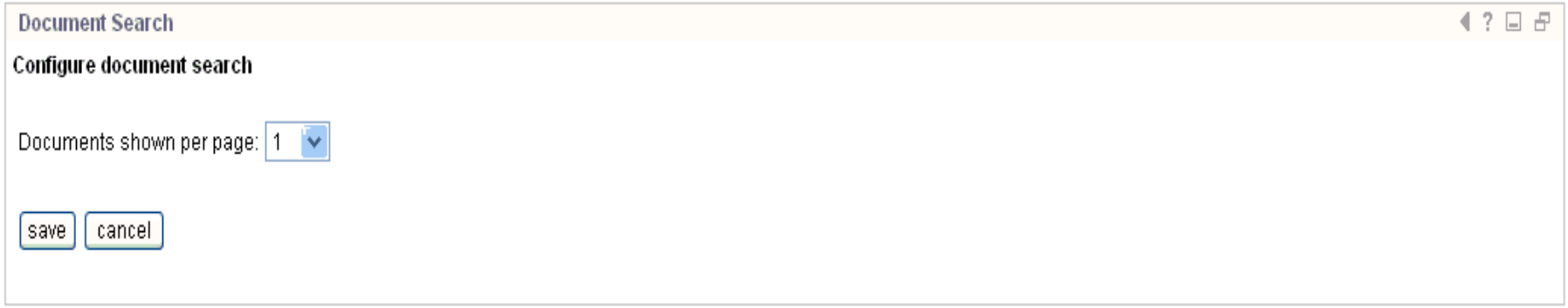
Manage search index



Perform a Search

- **Specify a search request, including some personal preferences:**
  - ►How many documents to present in the search results page
  - ►Ranking options
- **Execute the search**
- **View the search result list.**
- **Save user settings and apply to subsequent searches.**

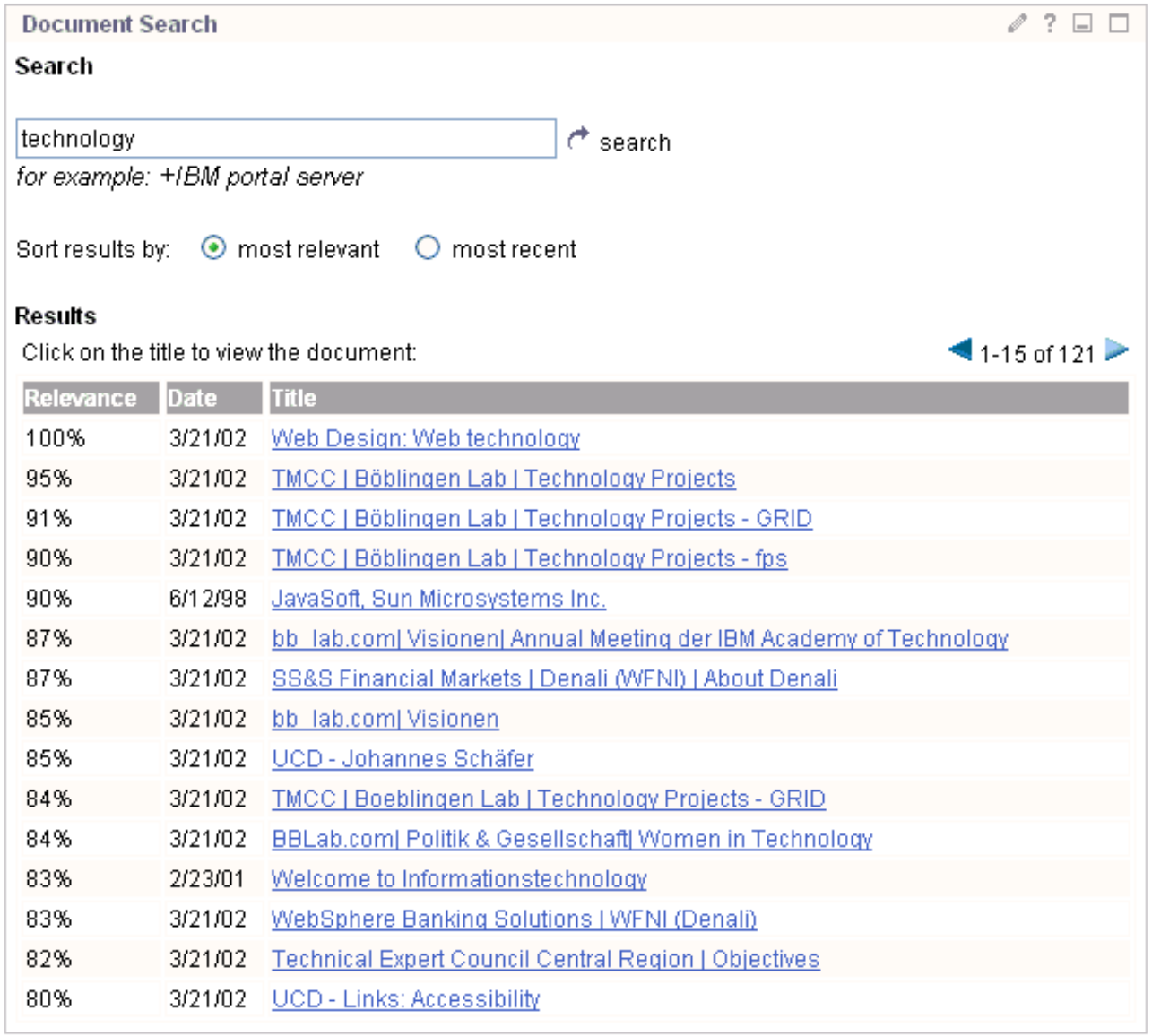
Search portlet - edit mode



User can choose to have 1, 2, 3, 4, 5, 10, 15, 20, 25 documents shown per result page. Juru always returns at maximum the top 200 documents.

Search Example

Search portlet - with result list



Relevance	Date	Title
100%	3/21/02	Web Design: Web technology
95%	3/21/02	TMCC | Böblingen Lab | Technology Projects
91%	3/21/02	TMCC | Böblingen Lab | Technology Projects - GRID
90%	3/21/02	TMCC | Böblingen Lab | Technology Projects - fps
90%	6/12/98	JavaSoft, Sun Microsystems Inc.
87%	3/21/02	bb_lab.com| Visionen| Annual Meeting der IBM Academy of Technology
87%	3/21/02	SS&S Financial Markets | Denali (WFNI) | About Denali
85%	3/21/02	bb_lab.com| Visionen
85%	3/21/02	UCD - Johannes Schäfer
84%	3/21/02	TMCC | Boeblingen Lab | Technology Projects - GRID
84%	3/21/02	BBLab.com| Politik & Gesellschaft| Women in Technology
83%	2/23/01	Welcome to Informationstechnology
83%	3/21/02	WebSphere Banking Solutions | WFNI (Denali)
82%	3/21/02	Technical Expert Council Central Region | Objectives
80%	3/21/02	UCD - Links: Accessibility

Domino Extended Search

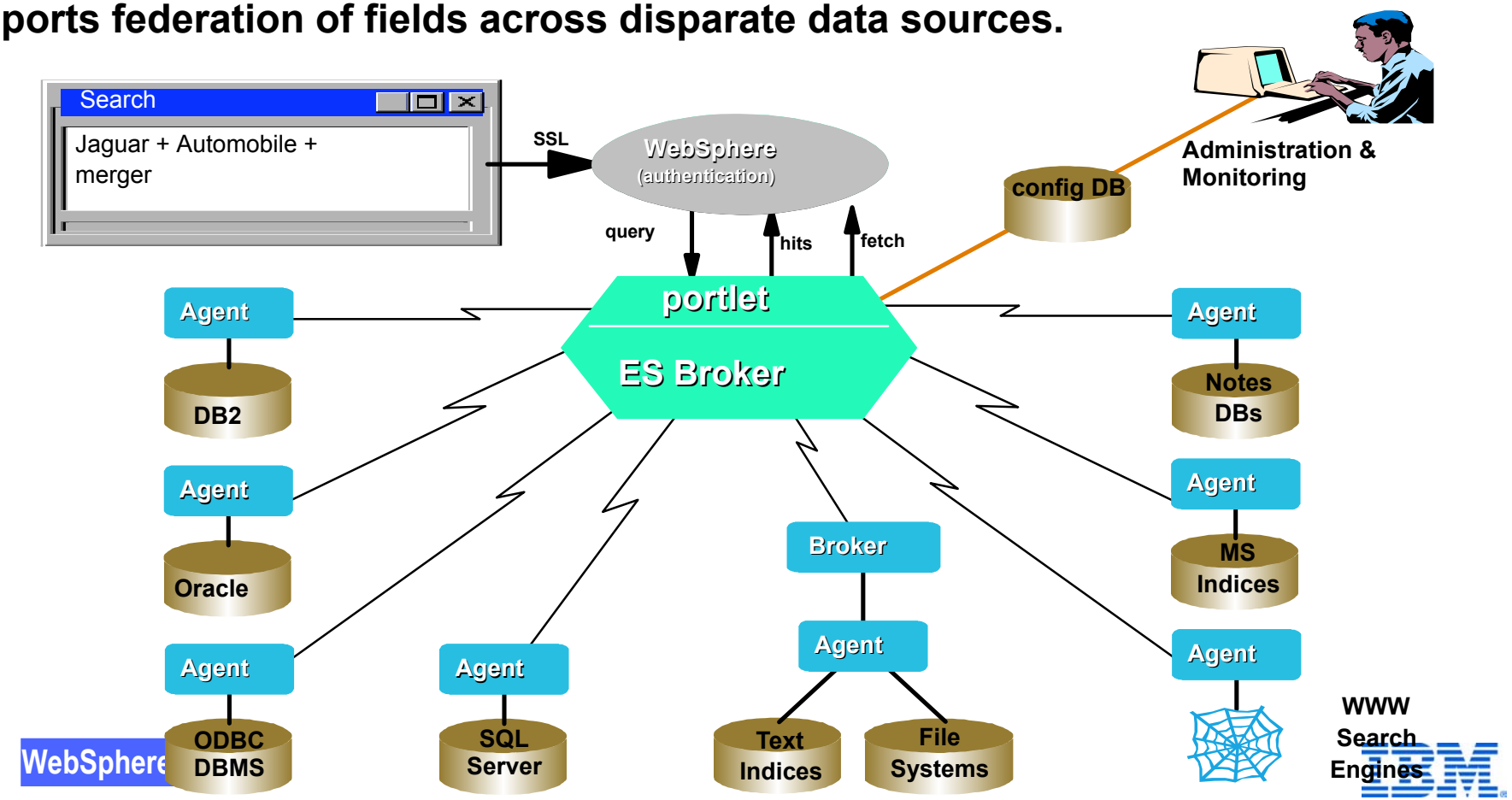
- **Domino Extended Search is a federated search engine that provides a single-point access to perform read-only, parallel, distributed, heterogeneous search capability across structured and unstructured data sources.**
  - Included with Extend and Experience offerings.
  - For clarification, Domino Extended Search, DES, IBM Extended Search, Extended Search, and ES all refer to the same product.
- **Extended Search portlet**

Other IBM Search Solutions

- **Two other federated search engines are available from IBM that will work with WebSphere Portal.**
  - Enterprise Information Portal (EIP)
  - Lotus Discovery Server.
  - Both are separately purchased products.

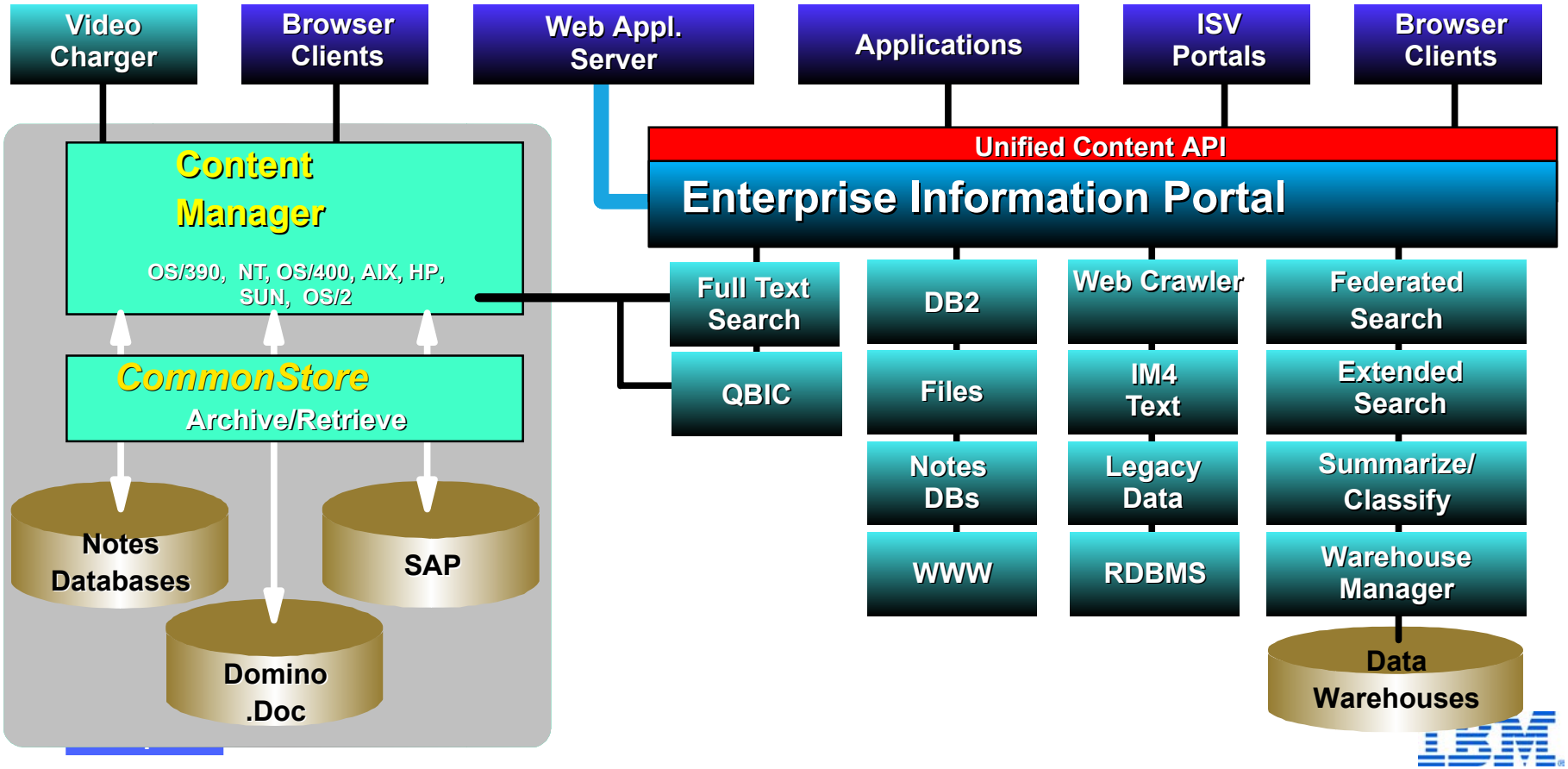
IBM Lotus Extended Search

- It accesses your information no matter where it is or how it is stored.
- It employs a distributed search technology.
- Supports multiple, heterogeneous indices.
- Does not require re-indexing of data.
- Lets indices be co-located or dispersed.
- Executes multiple searches in parallel.
- Supports federation of fields across disparate data sources.



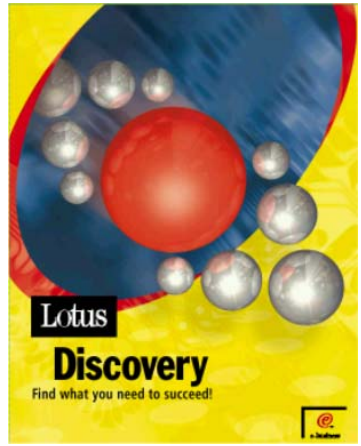
Enterprise Information Portal

- **IBM Enterprise Information Portal (EIP) is a set of tools for information integration**
  - ►A "Data Integrator" for data & information
  - ►Provides categorization, crawler, advanced federated search, integration with workflow, and many connectors to back-end repositories.



Lotus Discovery Server

- Visual Knowledge Discovery = Knowledge Map
- Experts, Documents, Places in single UI
- Find "Value, Context, Quality" of data
- Locate Experts
- Unprecedented Automation & Metrics Analysis
- Leverages Existing Core & KM Investments
- Can help "organize a mess" & Clean Up
- Knowledge Audits (Reports)
- Distributed processing
- Extensible (SDK, KDS API, LotusScript, etc.)



Search Comparisons

- **WebSphere Portal Search, IBM Lotus Extended Search, Lotus Discovery Server, Enterprise Information Portal**
  - WebSphere Portal Search: for indexing and searching web site's text
  - IBM Lotus Extended Search: for doing federated searches through multiple search engines.
  - Enterprise Information Portal: for doing federated searches of structured and unstructured data across entire enterprises.
  - Lotus Discovery Server: for searches combined with the ability to do knowledge and expert management.