Query-based Taxonomies

Next
THE BEST-RUN BUSINESSES RUN SAP
Exit
Bookmarks

Table of Contents

Introduction

Example-based Taxonomies

Query-based Taxonomies

Planning

THE BEST-RUN BUSINESSES RUN SAP

Multiple Physical Repositories

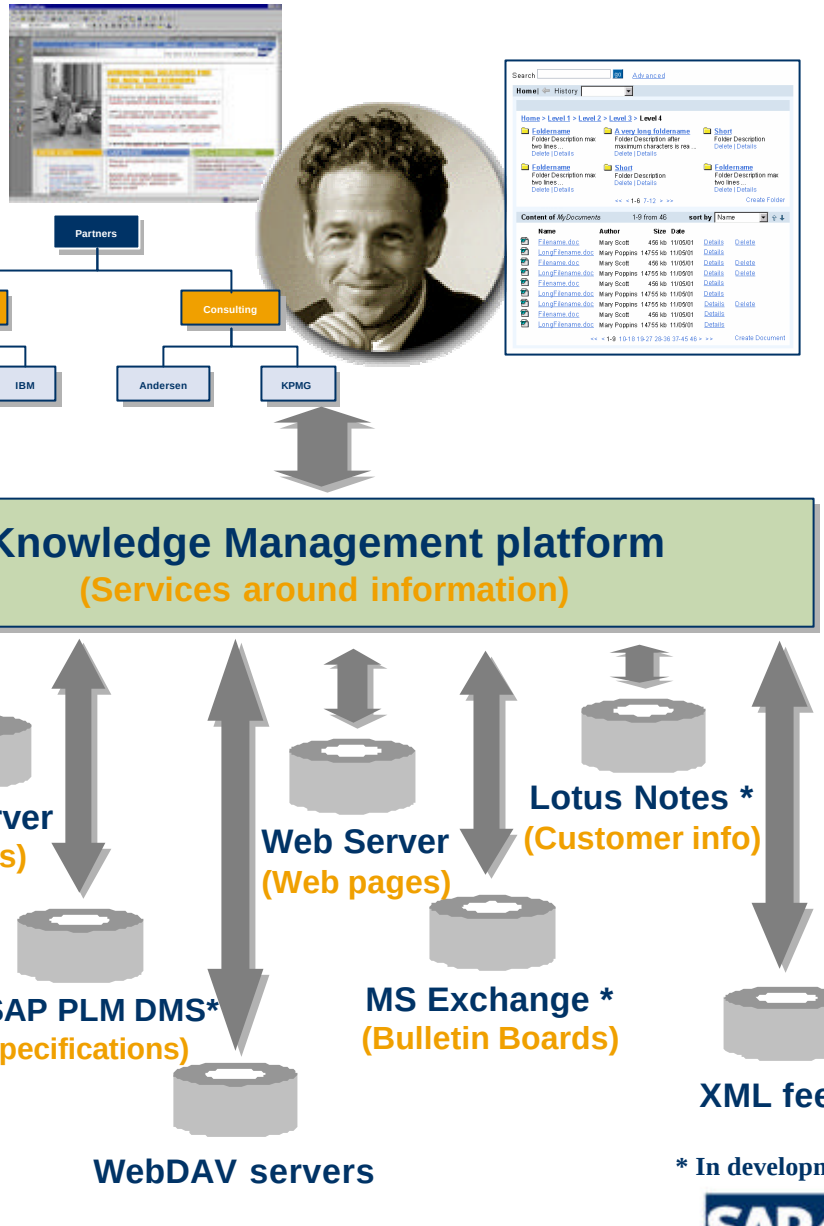
“It is impossible to centralize all the information in a single repository ...“

KM provides

- **A unified view, taxonomies and set of common services**
- **on top of heterogenous repositories of unstructured information**

KM consists of

- **Content Management**
- **Retrieval & Classification**
- **Collaboration**



THE BEST-RUN BUSINESSES RUN SAP

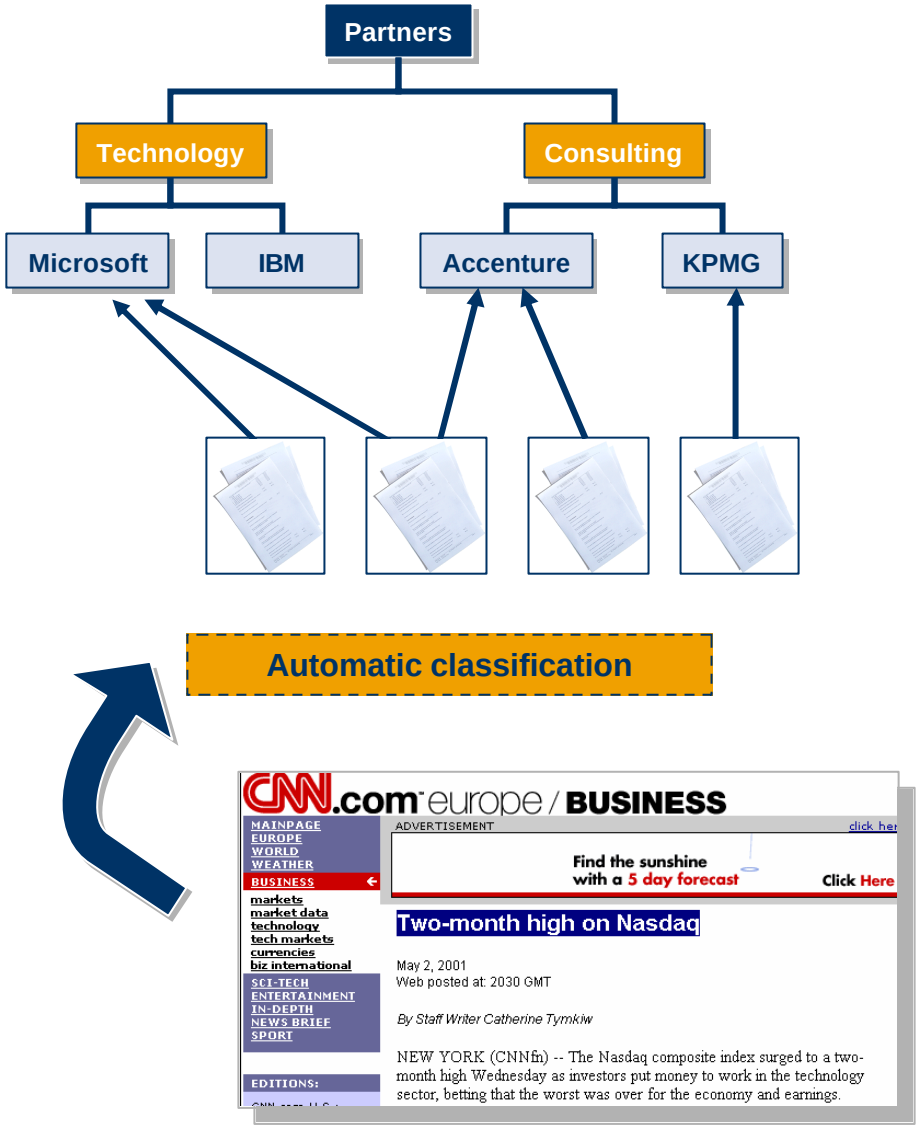
Virtual Navigation

Automatic Document Classification

- Information is organized in multiple taxonomies
- Full text search alone does not help
  - Known terms only
  - Terms may change over time

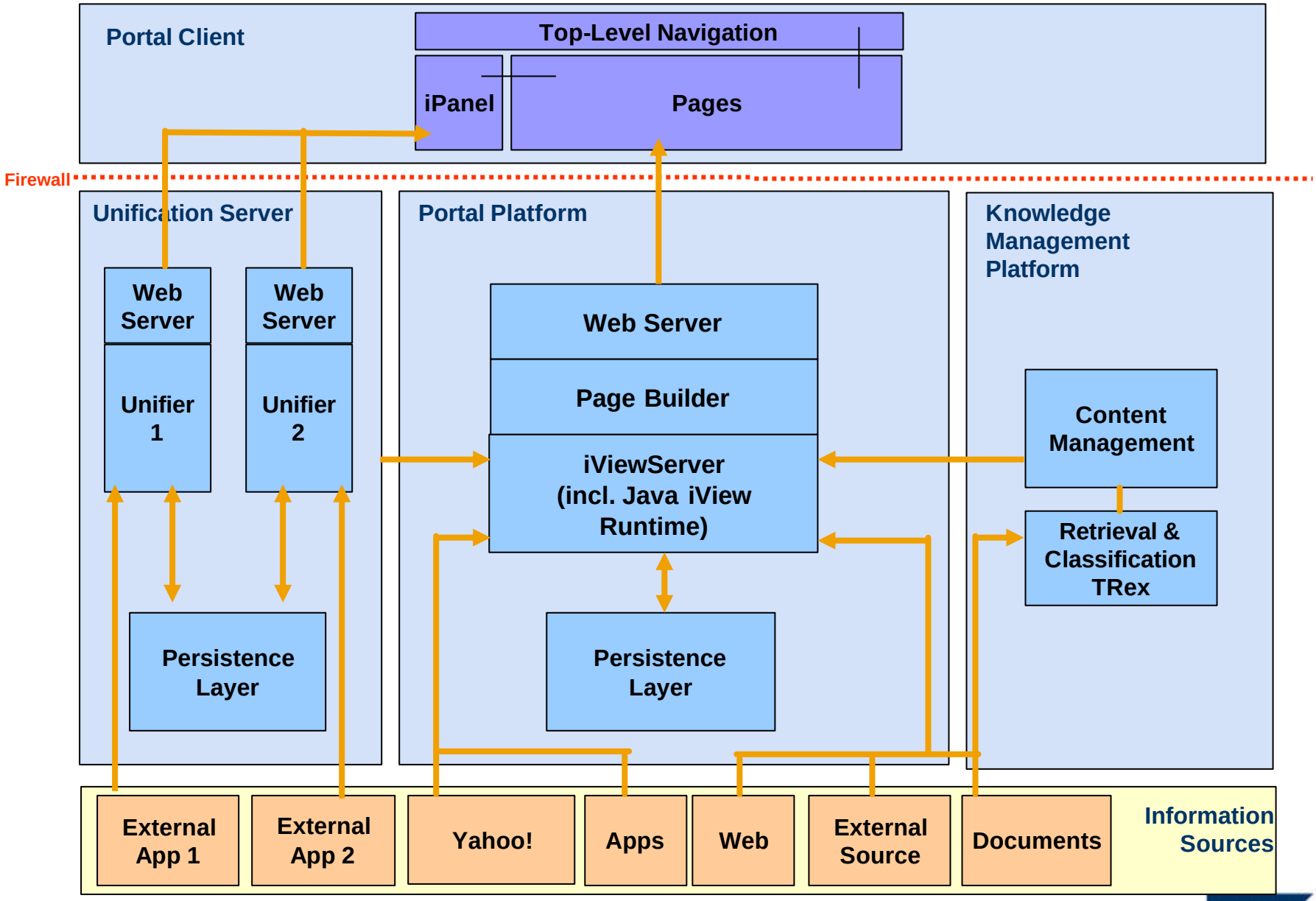
Classification Engine

- Self-learning algorithms or query-based definitions
- Unlimited taxonomies
- Vector model intelligence



THE BEST-RUN BUSINESSES RUN SAP

Involved Components



THE BEST-RUN BUSINESSES RUN SAP

Next
Back
Exit
Bookmarks

Table of Contents

Introduction

Example-based Taxonomies

Query-based Taxonomies

Planning

THE BEST-RUN BUSINESSES RUN SAP

Next

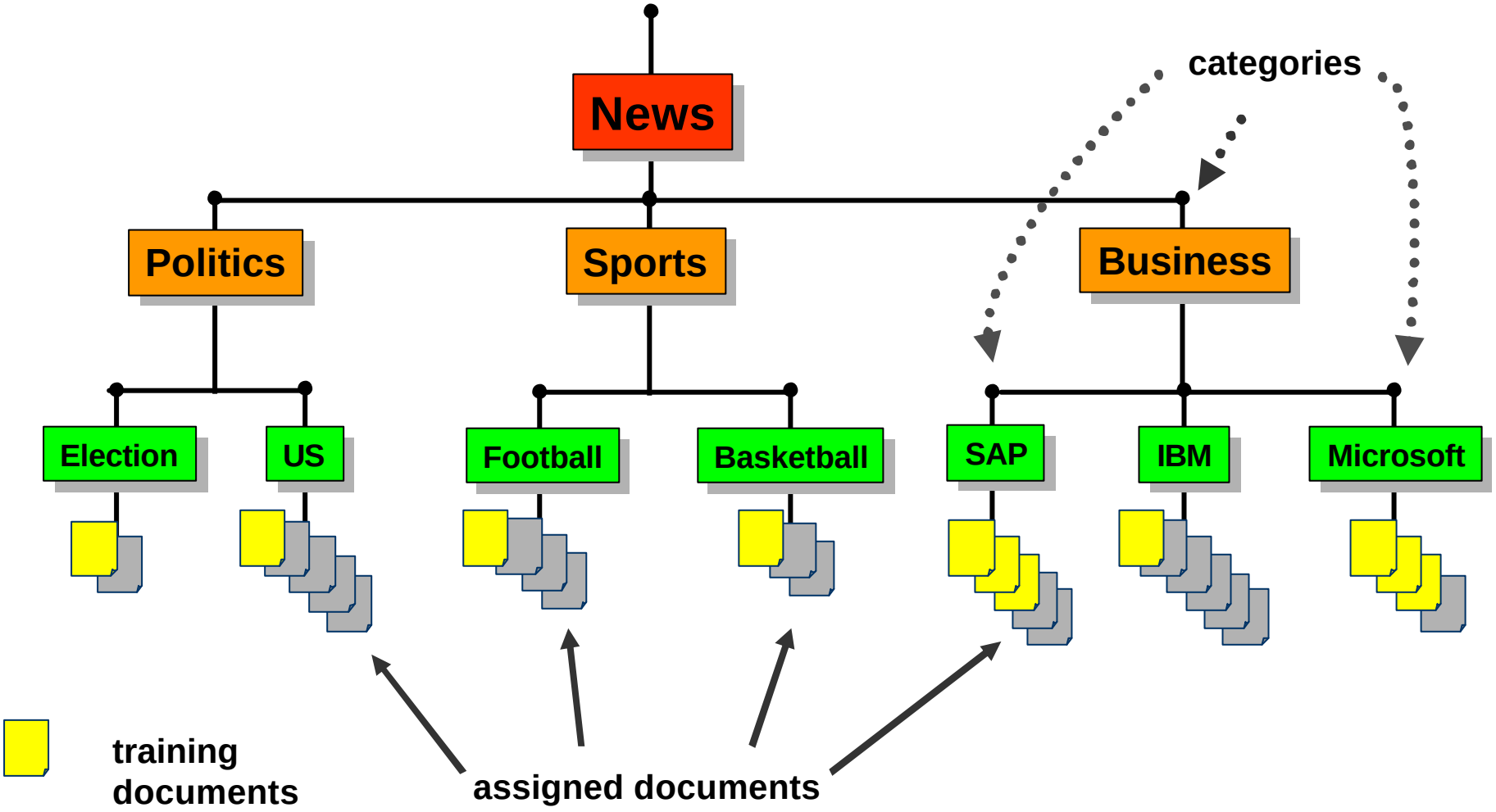
Back

Exit

Bookmarks

Concept : Example-based taxonomy

Since FCS of EP 5.0



THE BEST-RUN BUSINESSES RUN SAP

Next
Back
Exit
Bookmarks

Example-based Taxonomies

Issue to watch: Good sets of training documents

Categories are described by sample documents

- A bunch of sample documents for each category is needed as “trainings set”
- The classification results depend of the “quality” of the training sets
- A training set can be used for more than one taxonomy
- After setting up a training set -> automatic classification can be used for the attached documents folders

Advantage of example bases classification:

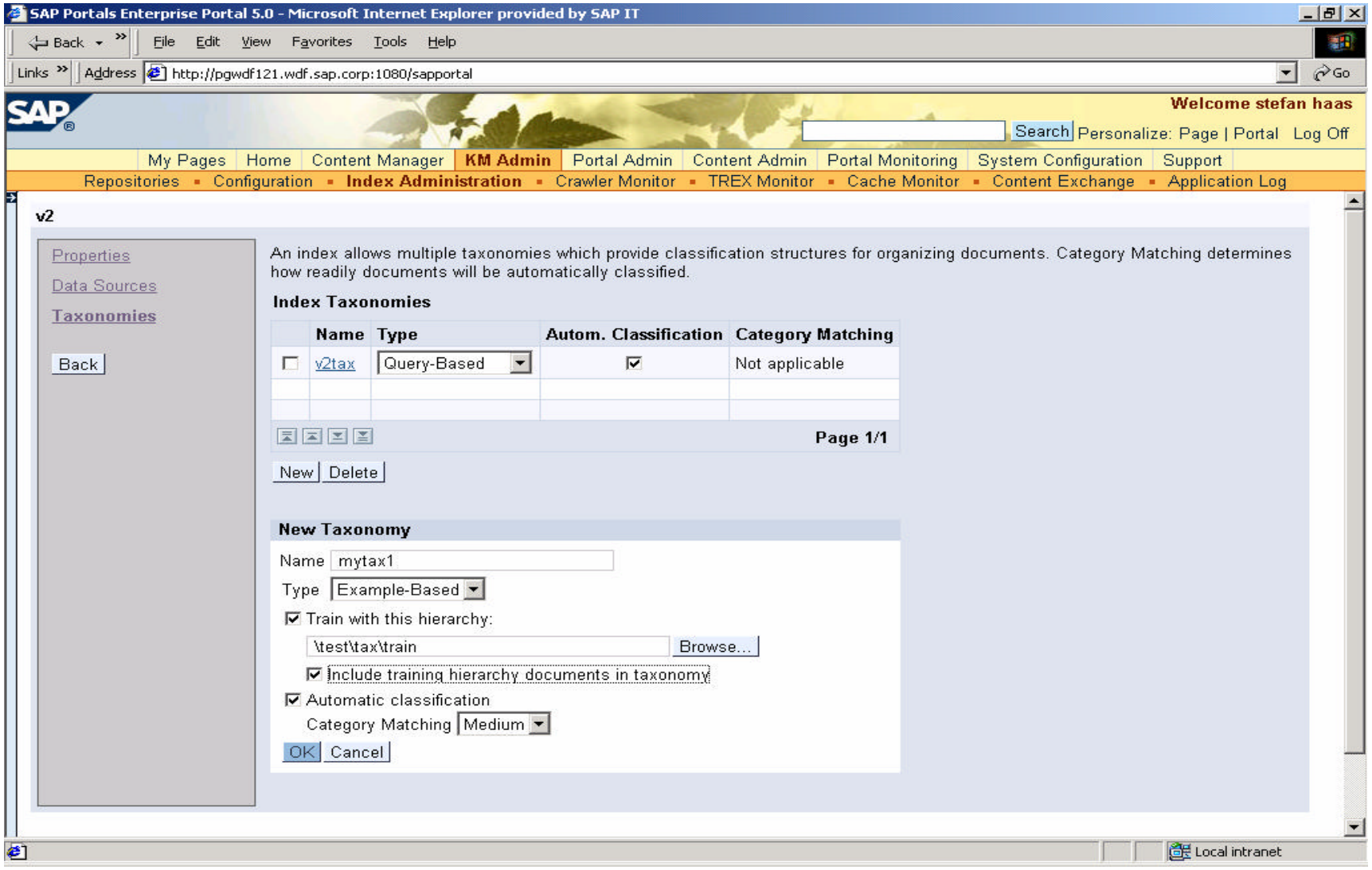
- Existing document structures can be used
- No manually maintained queries are needed

Disadvantage:

- Difficult verification of classification results
- Determining good sample documents

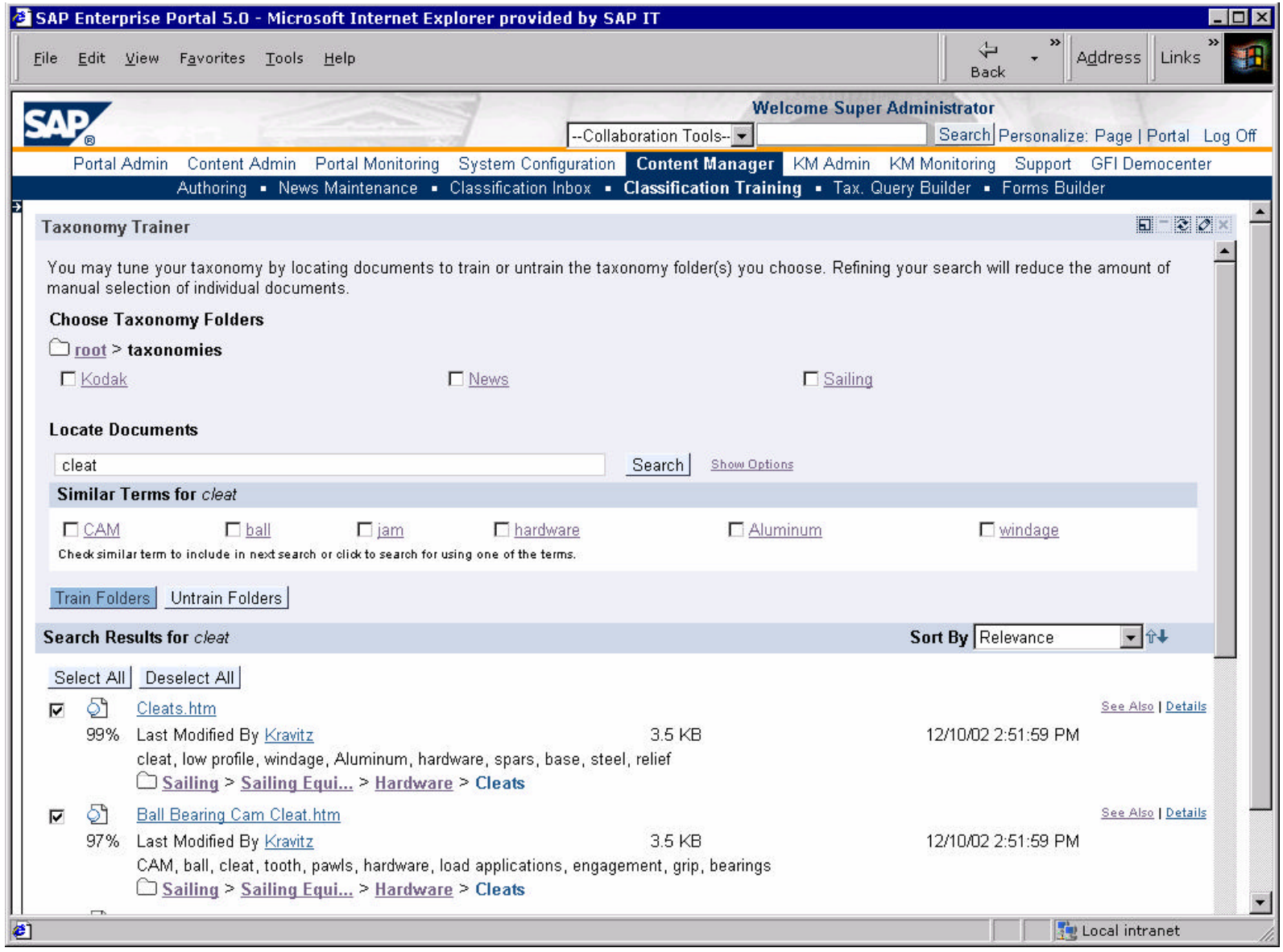
THE BEST-RUN BUSINESSES RUN SAP

Creating Example-based Taxonomies



THE BEST-RUN BUSINESSES RUN SAP

Train EBT with Search Results



THE BEST-RUN BUSINESSES RUN SAP

Table of Contents

Introduction

Example-based Taxonomies

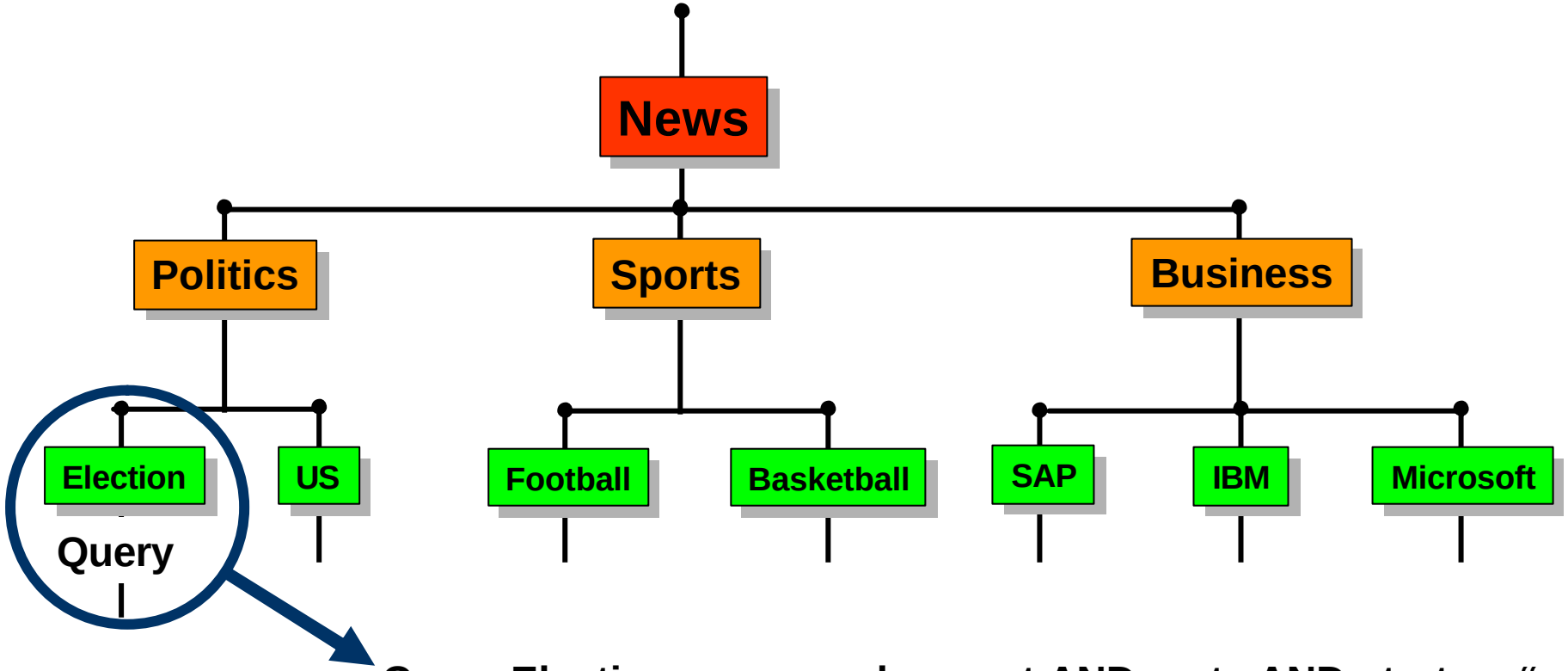
Query-based Taxonomies

Planning

THE BEST-RUN BUSINESSES RUN SAP

Concept : Query-based taxonomy I

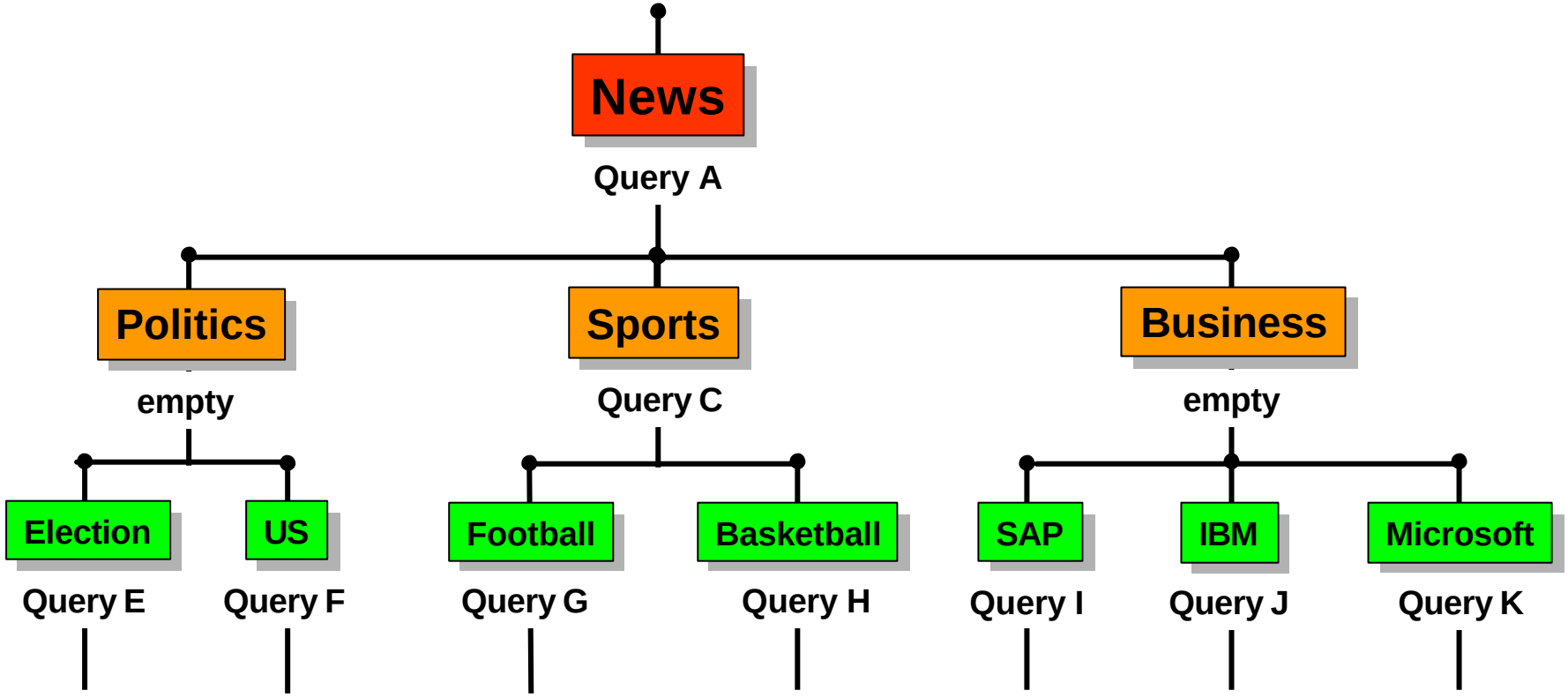
Since SP5 for EP 5.0



Query Election = „unemployment AND party AND strategy“

THE BEST-RUN BUSINESSES RUN SAP

Concept : Query-based Taxonomy II



Knowledge is one of the most important production factors. However, the lack of knowledge and information is constantly becoming more obvious. Information exists in numerous file formats that are stored using different kinds of media and in widely distributed storage locations. Because such information is often hidden in the darkest, most inaccessible corner of a company, it can be impossible for companies to retain an overview of these piles of information and to make the best use of their resources.

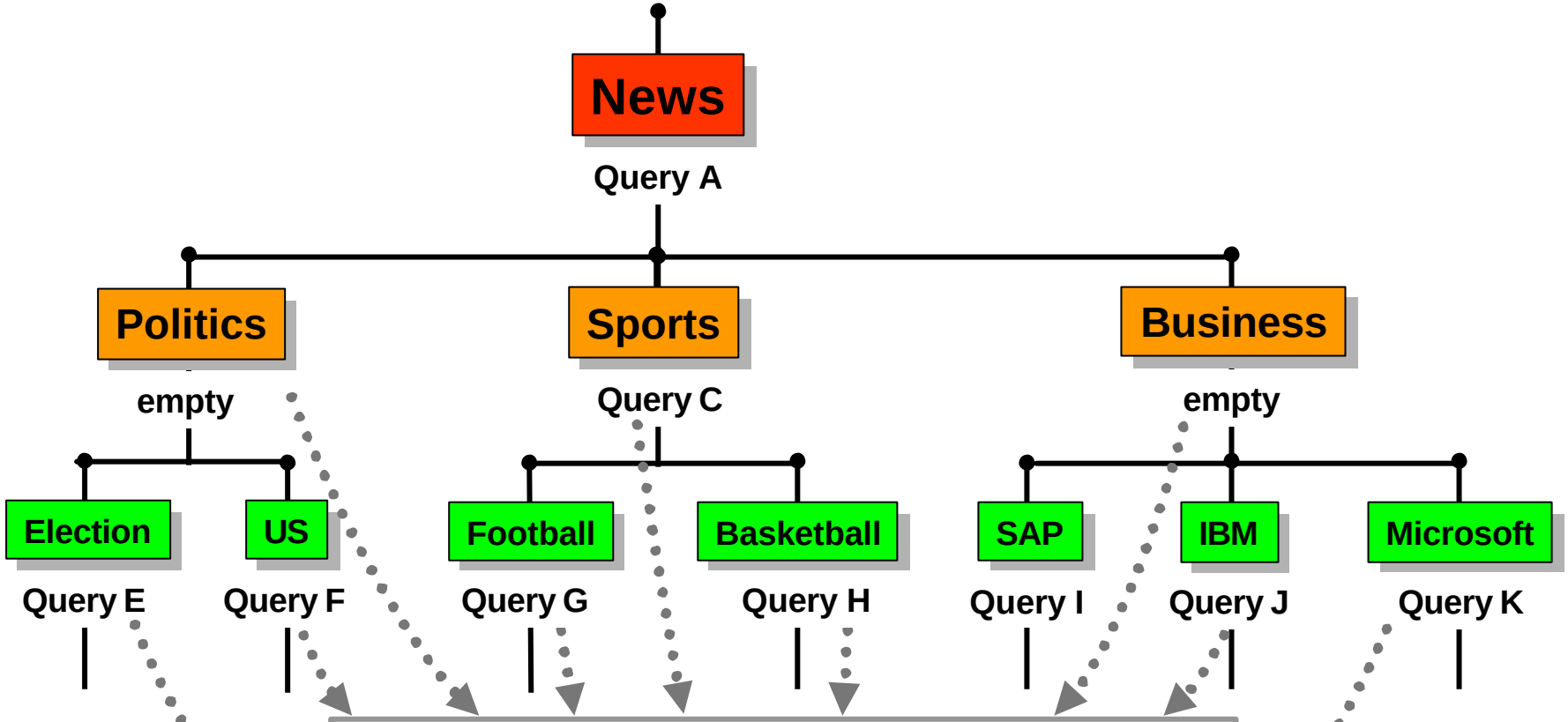
If query matches: list of classes

DOCUMENT

THE BEST-RUN BUSINESSES RUN SAP

Next
Back
Exit
Bookmarks

Concept : Query-based Taxonomy II



THE BEST-RUN BUSINESSES RUN SAP

Concept : Query Based Taxonomies III

Issue to watch: Queries must discriminate, but should not be too detailed

Set a query on a category

- Boolean operators
- Searching content and attributes values

Categories

- Categories build a hierarchy
- A hierarchy can be modified (e.g. add a category, delete a category)
- Categories with no queries will be ignored
- Can be moved to another parent

Queries

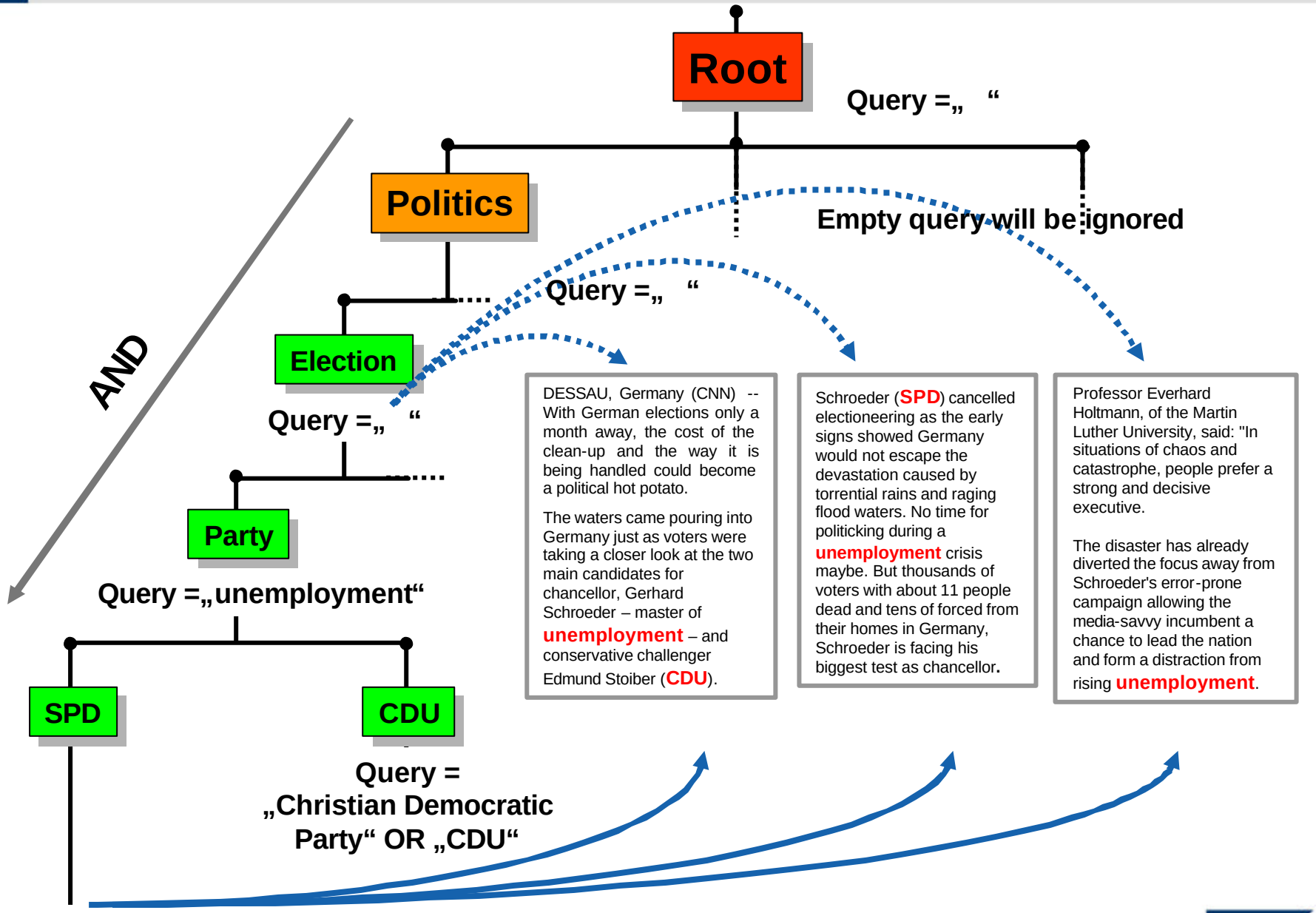
- are hierarchical
- Modification of queries is possible

Query based classification :

- Queries will be fired against a given document
  if the query matches -> the category will be given back as
  a classification proposal

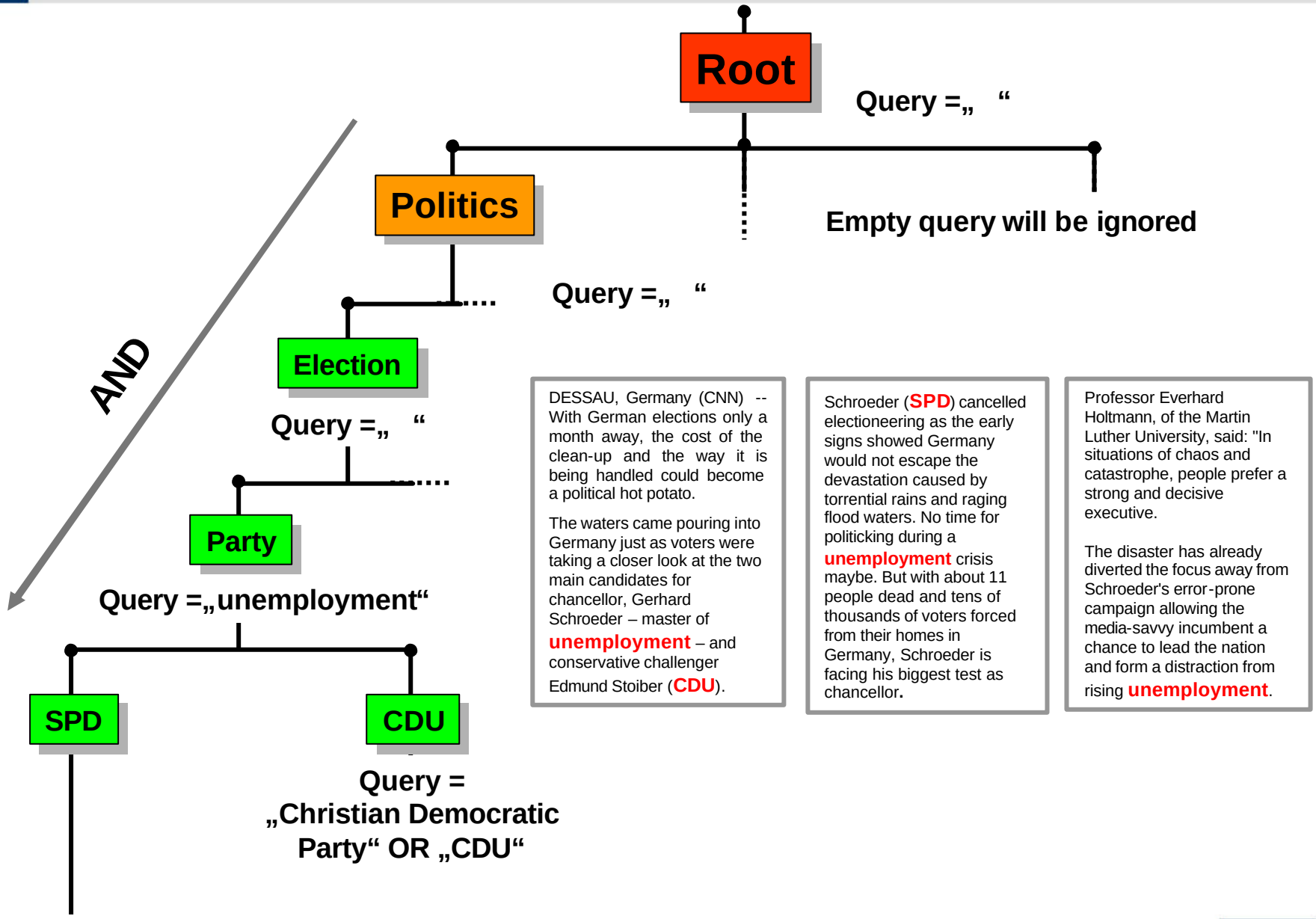
THE BEST-RUN BUSINESSES RUN SAP

Hierarchical Queries I



THE BEST-RUN BUSINESSES RUN SAP

Hierarchical Queries II



DESSAU, Germany (CNN) -- With German elections only a month away, the cost of the clean-up and the way it is being handled could become a political hot potato.

The waters came pouring into Germany just as voters were taking a closer look at the two main candidates for chancellor, Gerhard Schroeder – master of **unemployment** – and conservative challenger Edmund Stoiber (**CDU**).

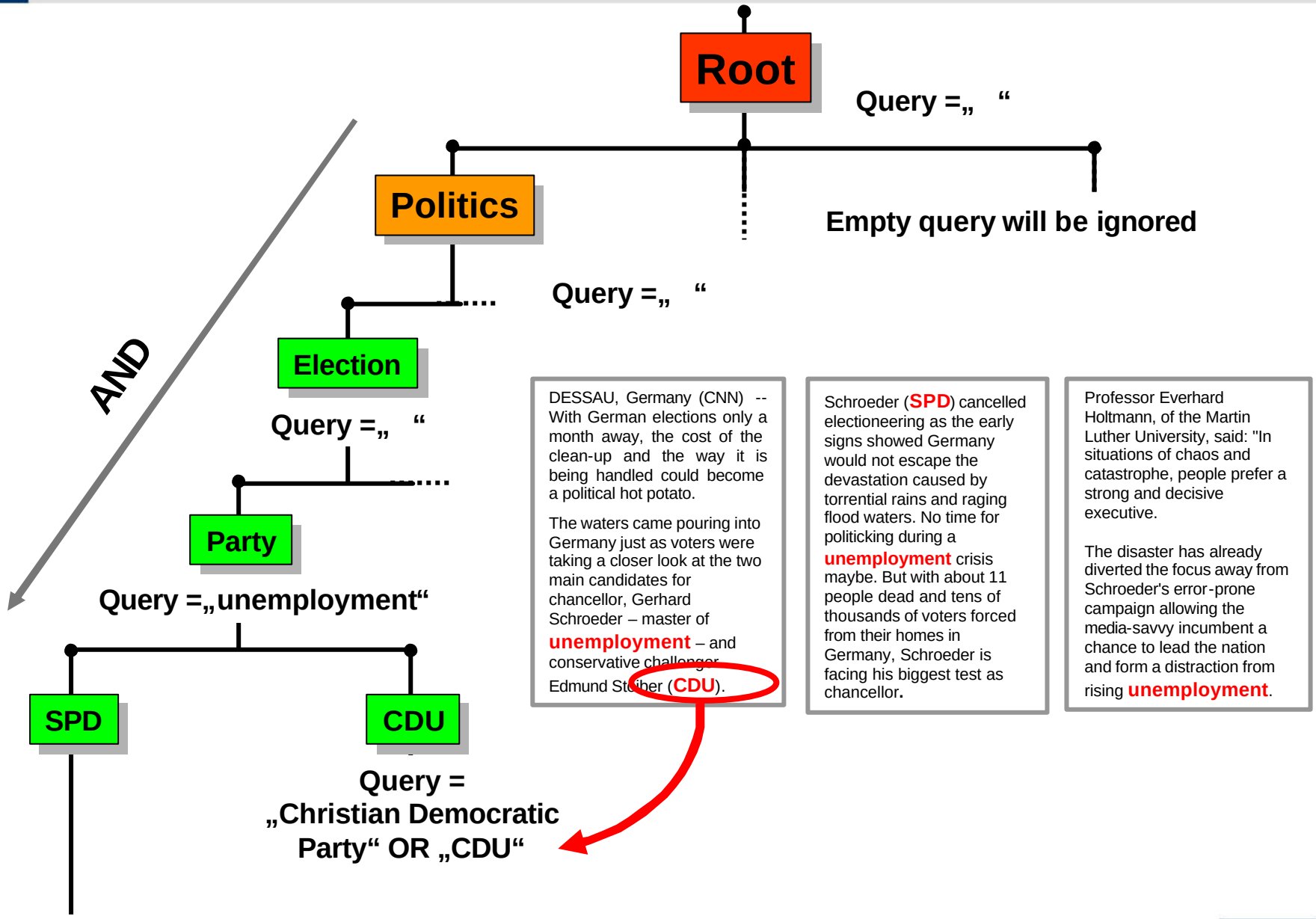
Schroeder (**SPD**) cancelled electioneering as the early signs showed Germany would not escape the devastation caused by torrential rains and raging flood waters. No time for politicking during a **unemployment** crisis maybe. But with about 11 people dead and tens of thousands of voters forced from their homes in Germany, Schroeder is facing his biggest test as chancellor.

Professor Everhard Holtmann, of the Martin Luther University, said: "In situations of chaos and catastrophe, people prefer a strong and decisive executive.

The disaster has already diverted the focus away from Schroeder's error-prone campaign allowing the media-savvy incumbent a chance to lead the nation and form a distraction from rising **unemployment**.

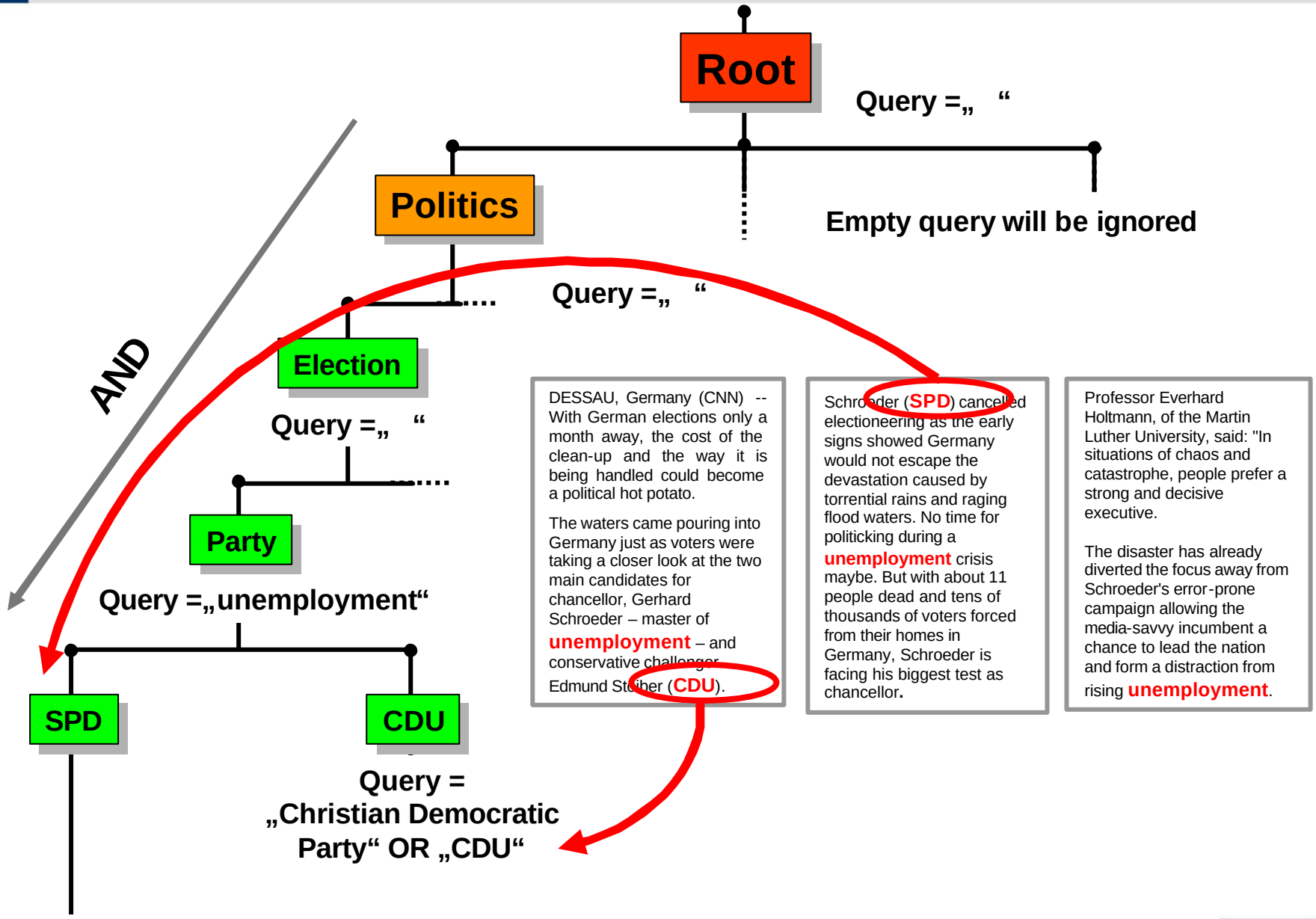
THE BEST-RUN BUSINESSES RUN SAP

Hierarchical Queries II



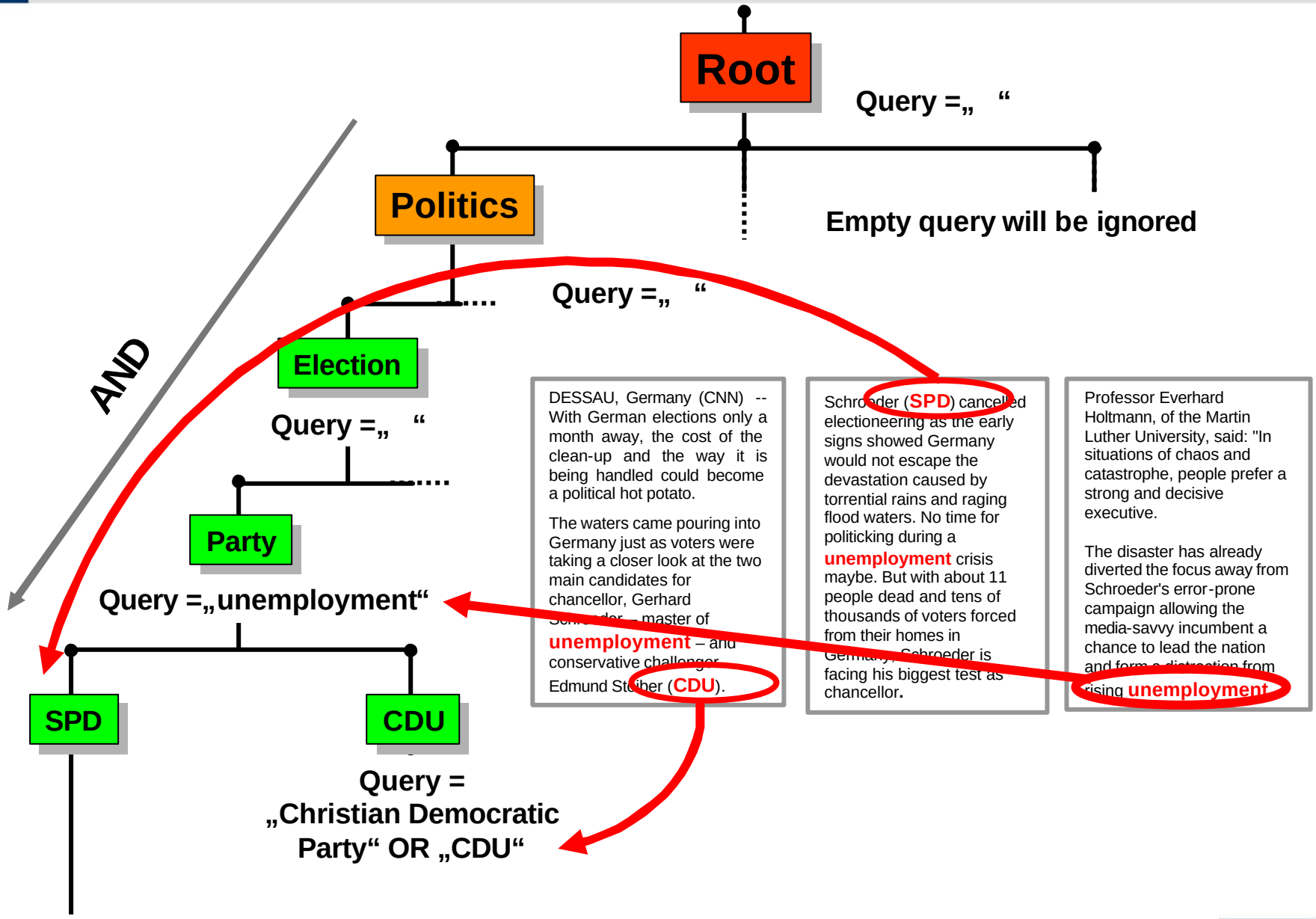
THE BEST-RUN BUSINESSES RUN SAP

Hierarchical Queries II



THE BEST-RUN BUSINESSES RUN SAP

Hierarchical Queries II



THE BEST-RUN BUSINESSES RUN SAP

Next
Back
Exit
Bookmarks

Hierarchical Queries III

- A document should belong to the most specific category
- All Queries from the root category must match -> “AND” connection from one level to next
- Empty queries are ignored (all items match)
- If the query of a child category will not match -> the parent category will be the suggestion
- Items can be assigned to more than one category

THE BEST-RUN BUSINESSES RUN SAP

Initial Setup and Maintenance

1. **Initial query based classification**
   - **a) create an index and attach a folder to an index**
   - **b) index the documents**
   - **c) define the taxonomy and the queries**
   - **d) update classification for all documents in the index**

-> advantage : better performance because mass document classification

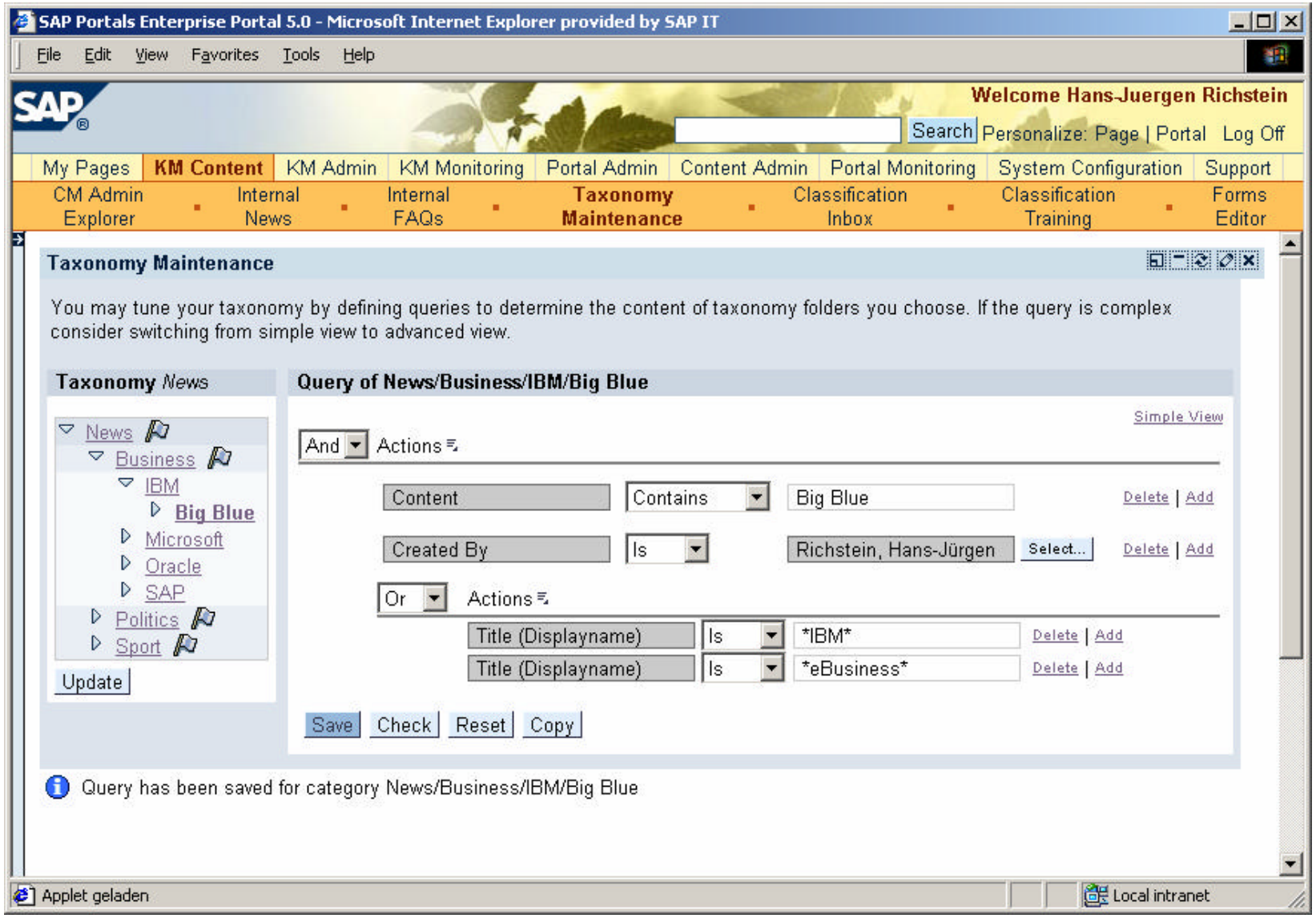
2. **Initial query based classification**
   - **a) create an index**
   - **b) define the taxonomy and the queries**
   - **c) attach a folder to an index**
   - **d) (automatic) classification for all documents (if auto classify flag is set)**
   - **e) index the documents**

-> disadvantage : poor performance because single document classification

3 a) Automatic classification of new documents
3 b) Manual handling of single documents in the Classification Inbox
3 c) Manual handling of single documents via “details” dialog

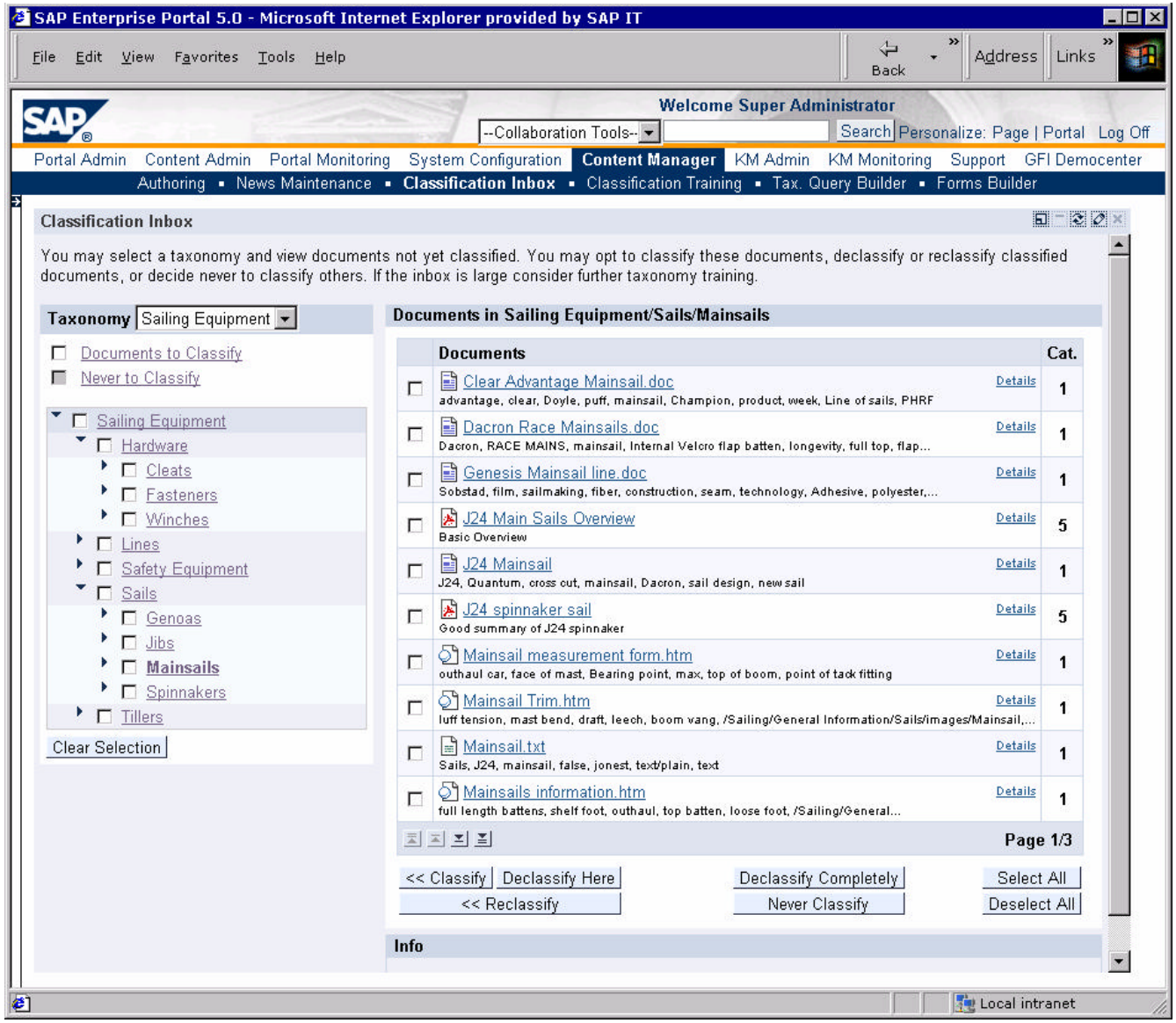
THE BEST-RUN BUSINESSES RUN SAP

Taxonomy Maintenance



THE BEST-RUN BUSINESSES RUN SAP

Classification Inbox for any Taxonomy



THE BEST-RUN BUSINESSES RUN SAP

Table of Contents

Introduction

Example-based Taxonomies

Query-based Taxonomies

Planning

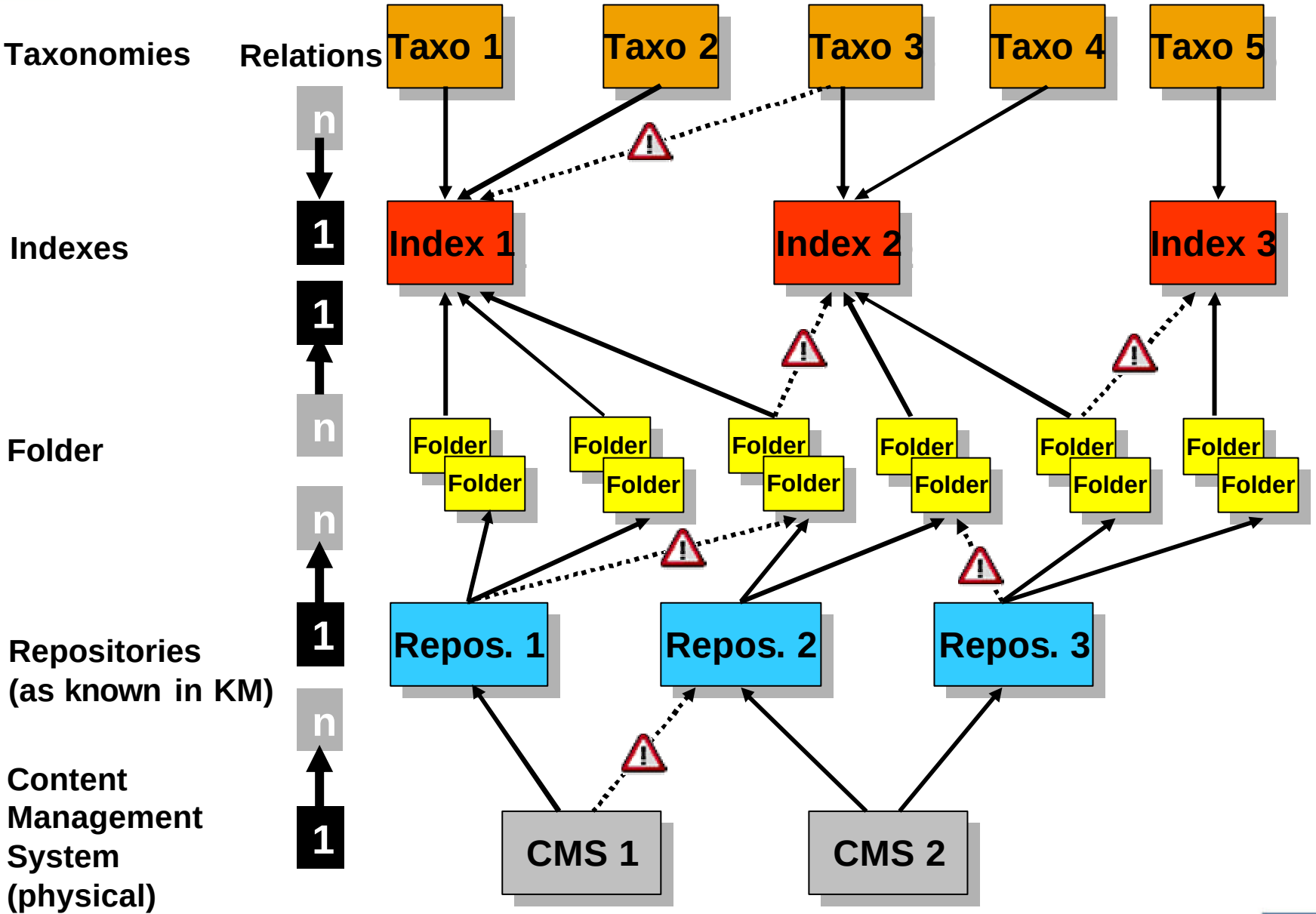
THE BEST-RUN BUSINESSES RUN SAP

Planning

- **Combination of example- and query-based classification**
- **Using class features (example based taxonomies) for the definition of queries**
- **Remove the restrictions on the following slide...**

THE BEST-RUN BUSINESSES RUN SAP

Repositories, Folders, Indices, Taxonomies



THE BEST-RUN BUSINESSES RUN SAP

Back
Exit
Bookmarks