

Query-based Taxonomies

Table of Contents



Introduction

Example-based Taxonomies

Query-based Taxonomies

Planning



Next



Back



Exit



Bookmarks

Multiple Physical Repositories

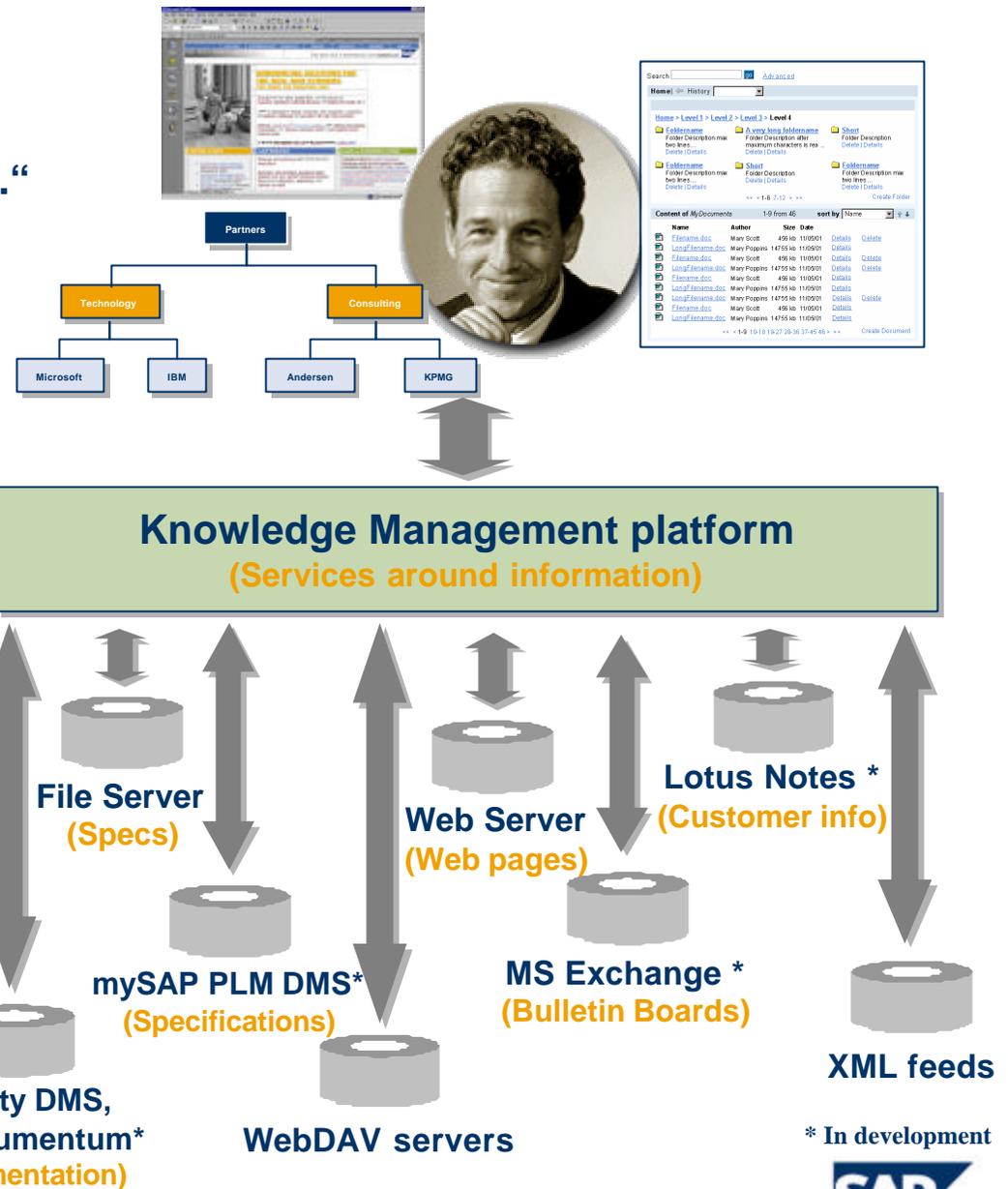
“It is impossible to centralize all the information in a single repository ...”

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



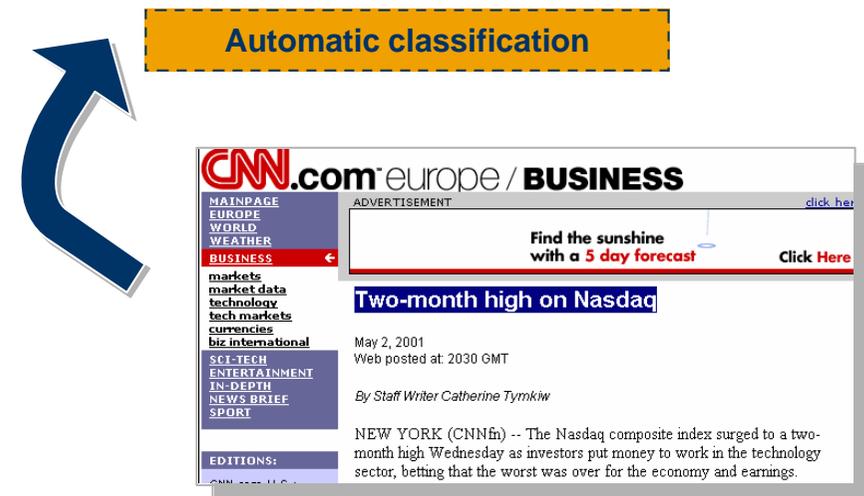
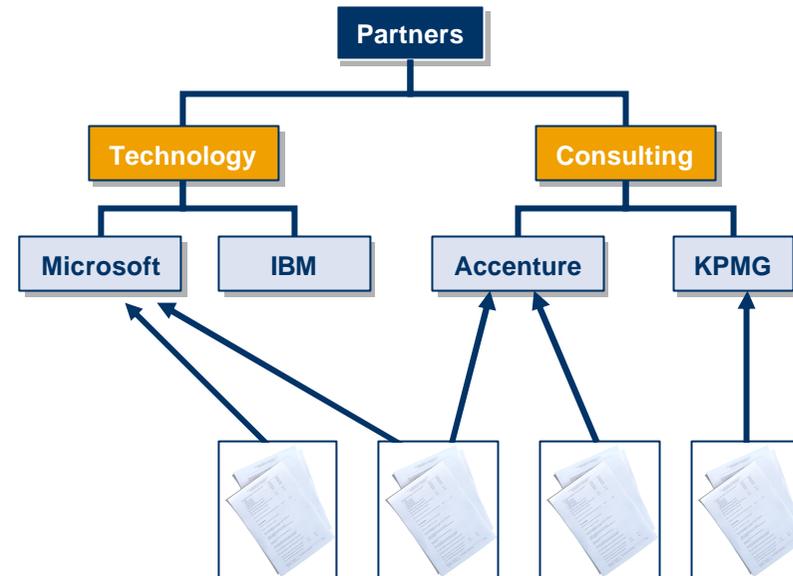
Virtual Navigation

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



Involved Components

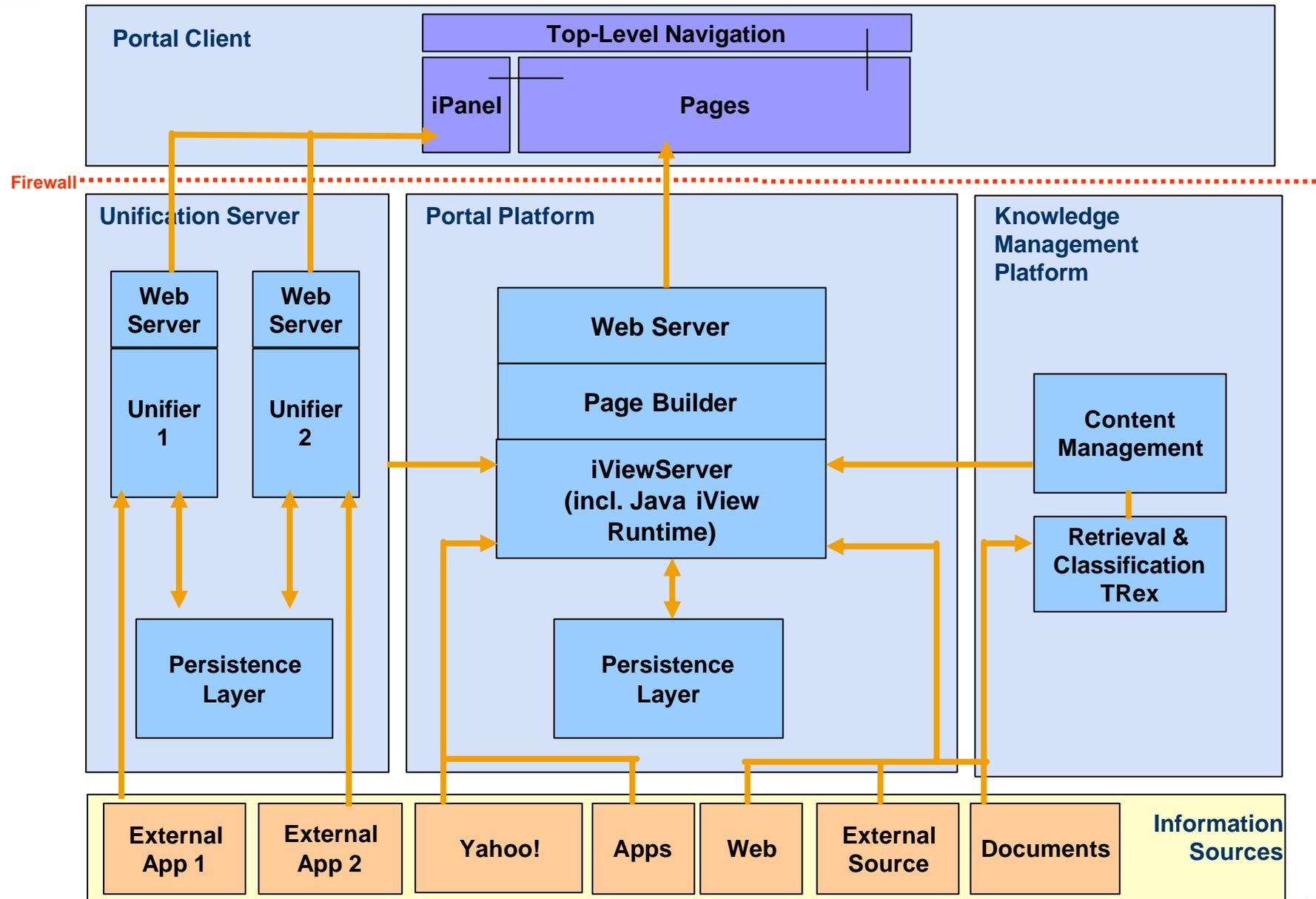


Table of Contents



Introduction

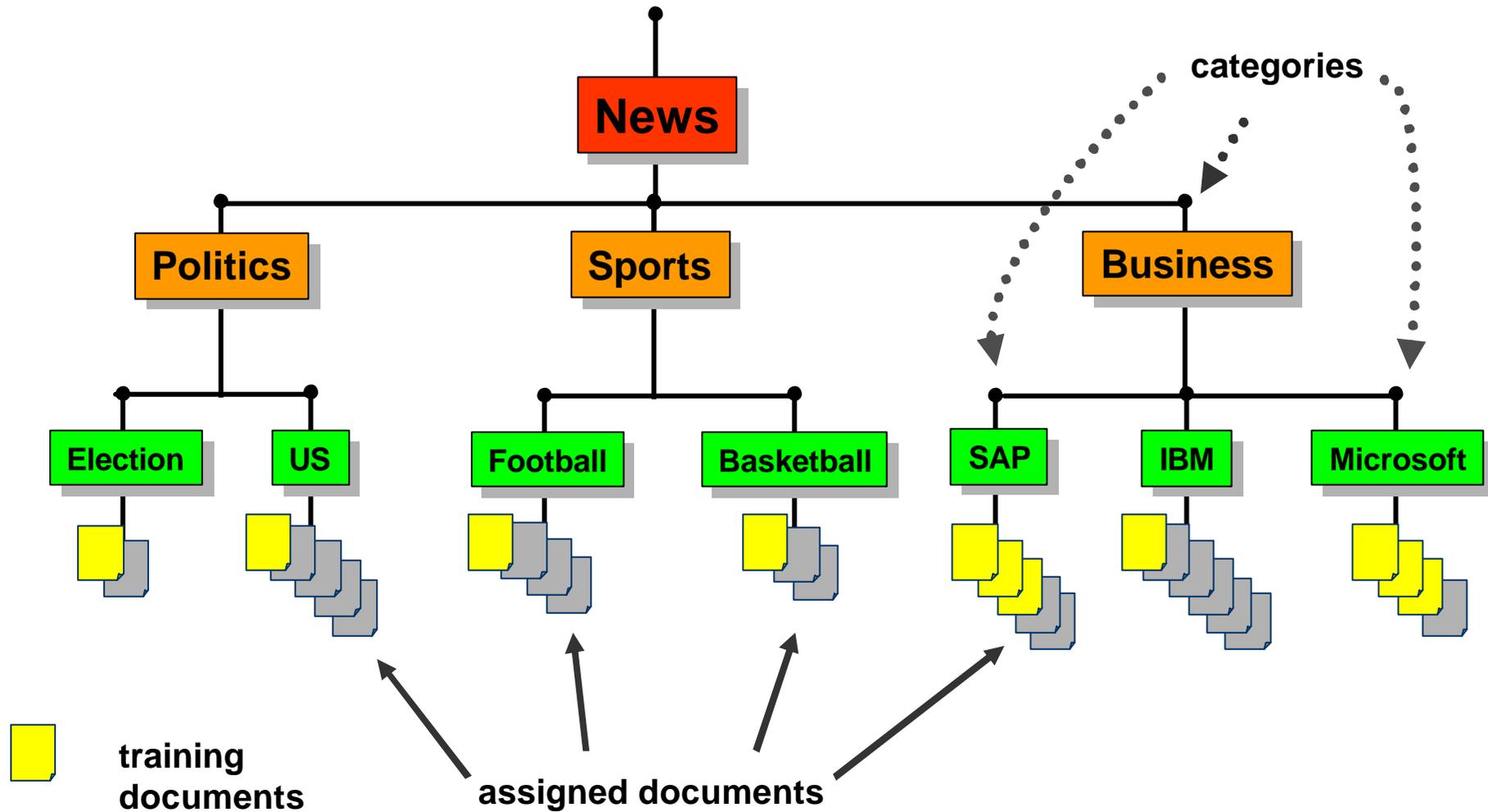
Example-based Taxonomies

Query-based Taxonomies

Planning

Concept : Example-based taxonomy

Since FCS of EP 5.0



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

Creating Example-based Taxonomies

SAP Portals Enterprise Portal 5.0 - Microsoft Internet Explorer provided by SAP IT

Address: http://pgwdf121.wdf.sap.corp:1080/sapportal

Welcome stefan haas

Search Personalize: Page | Portal Log Off

My Pages Home Content Manager **KM Admin** Portal Admin Content Admin Portal Monitoring System Configuration Support

Repositories Configuration **Index Administration** Crawler Monitor TREX Monitor Cache Monitor Content Exchange Application Log

v2

Properties
Data Sources
Taxonomies
Back

An index allows multiple taxonomies which provide classification structures for organizing documents. Category Matching determines how readily documents will be automatically classified.

Index Taxonomies

Name	Type	Autom. Classification	Category Matching
<input type="checkbox"/> v2tax	Query-Based	<input checked="" type="checkbox"/>	Not applicable

Page 1/1

New Delete

New Taxonomy

Name

Type

Train with this hierarchy:

include training hierarchy documents in taxonomy

Automatic classification
Category Matching

Train EBT with Search Results

The screenshot shows the SAP Enterprise Portal 5.0 interface in Microsoft Internet Explorer. The browser title is "SAP Enterprise Portal 5.0 - Microsoft Internet Explorer provided by SAP IT". The page header includes the SAP logo, a search bar, and navigation links like "Personalize: Page | Portal | Log Off". The main navigation bar contains "Content Manager" and other modules. The "Taxonomy Trainer" section is active, showing instructions on how to tune taxonomy. It lists "Choose Taxonomy Folders" with options for "Kodak", "News", and "Sailing". Below this is a "Locate Documents" section with a search input containing "cleat" and a "Search" button. A "Similar Terms for cleat" section lists terms like "CAM", "ball", "jam", "hardware", "Aluminum", and "windage". At the bottom, "Search Results for cleat" are displayed, sorted by "Relevance". Two results are shown, both checked for selection:

File Name	Relevance	Last Modified By	Size	Date
Cleats.htm	99%	Kravitz	3.5 KB	12/10/02 2:51:59 PM
Ball Bearing Cam Cleat.htm	97%	Kravitz	3.5 KB	12/10/02 2:51:59 PM

Table of Contents



Introduction

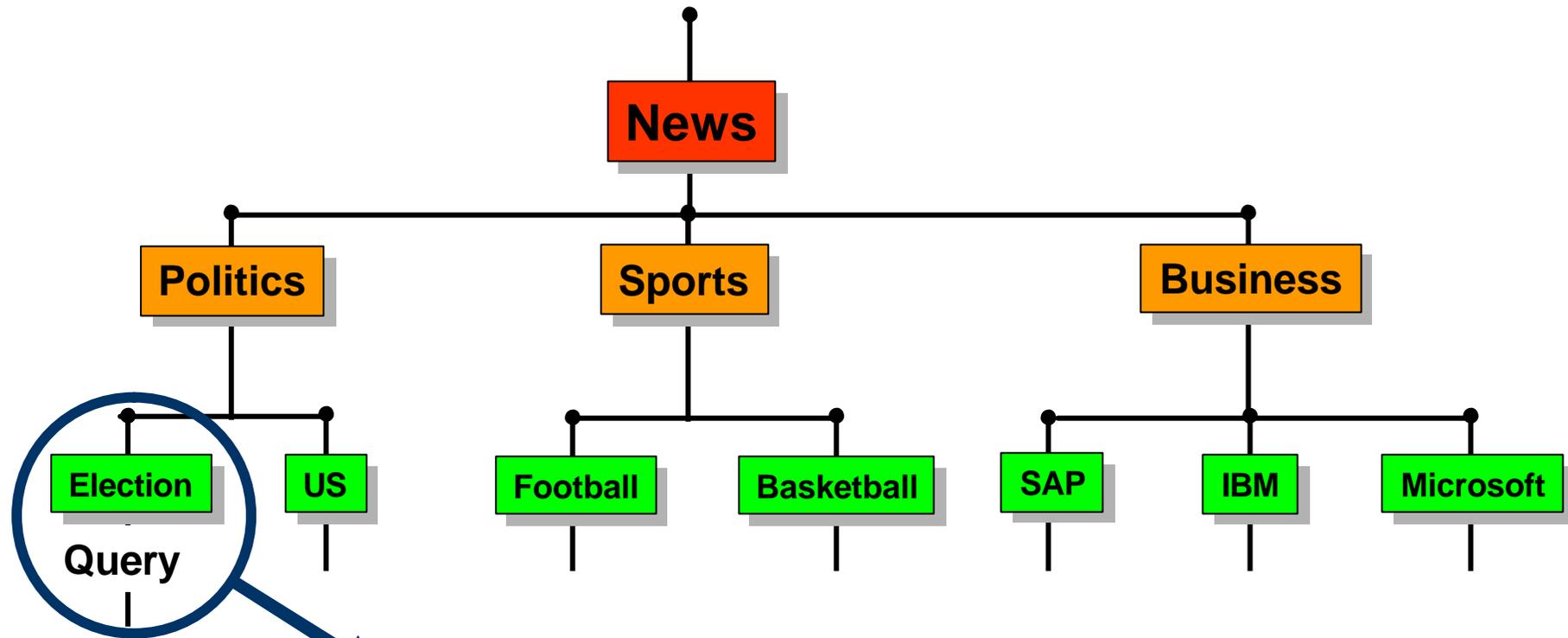
Example-based Taxonomies

Query-based Taxonomies

Planning

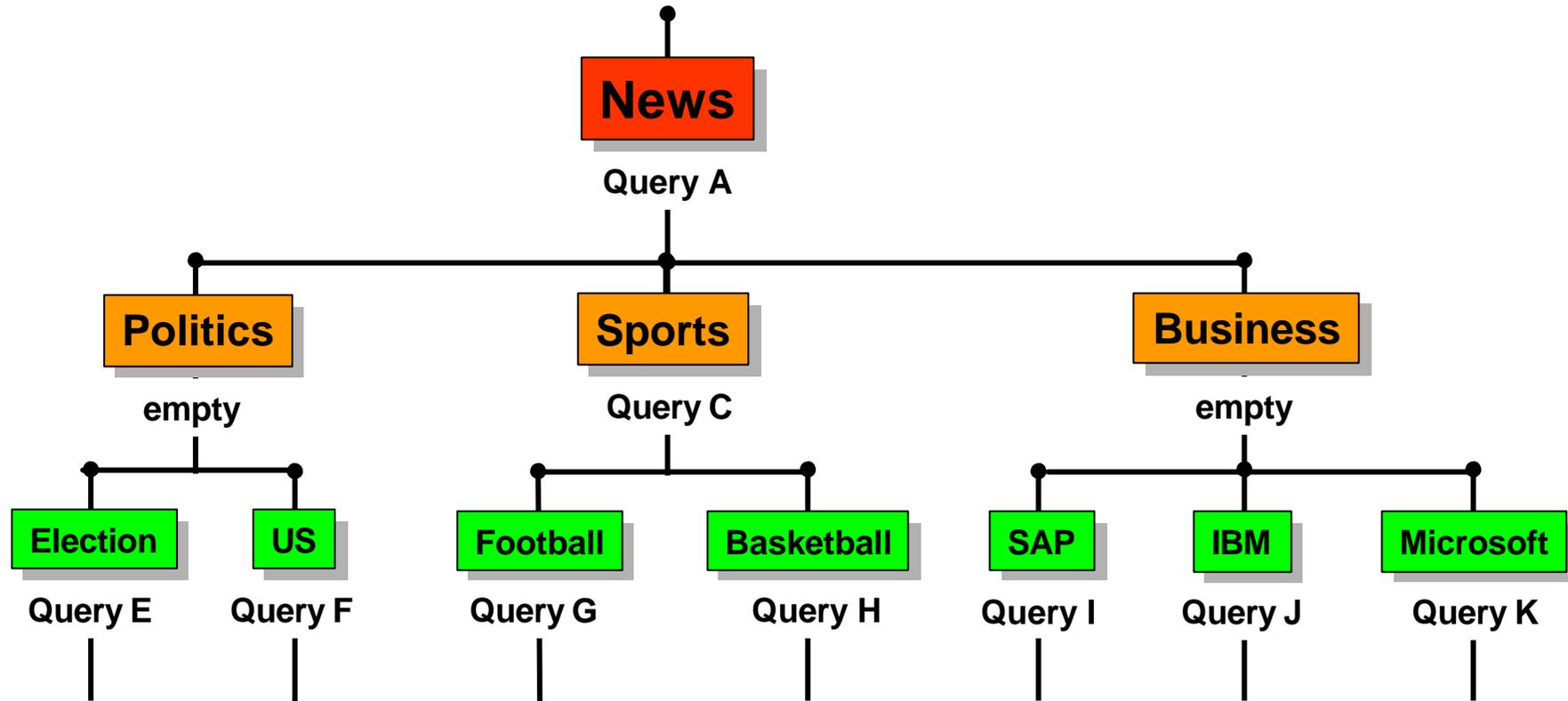
Concept : Query-based taxonomy I

Since SP5 for EP 5.0



Query Election = „unemployment AND party AND strategy“

Concept : Query-based Taxonomy II

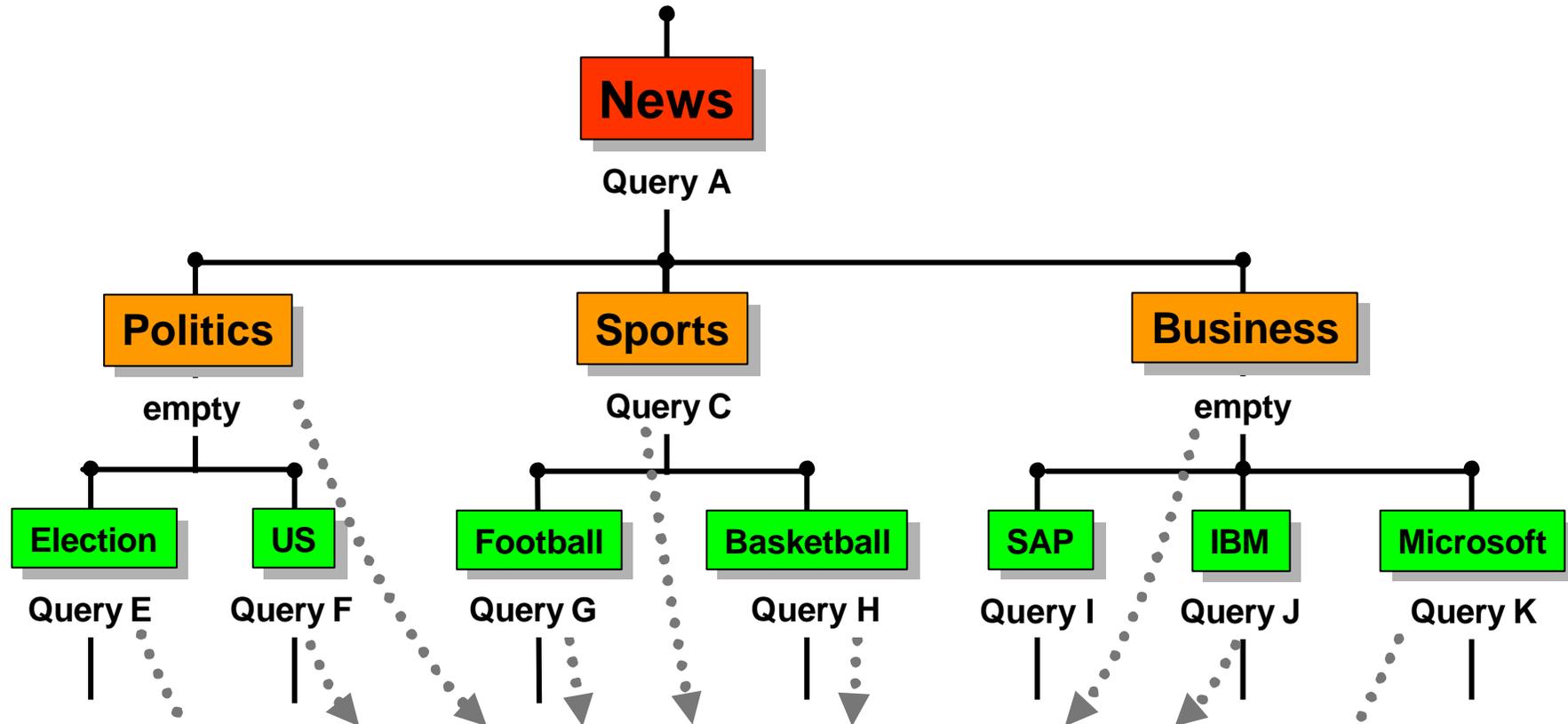


Knowledge is one of the most important production factors. However, the lack of knowledge and information is constantly becoming more obvious. Information exists in numerous file formats that are stored using different kinds of media and in widely distributed storage locations. Because such information 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

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

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

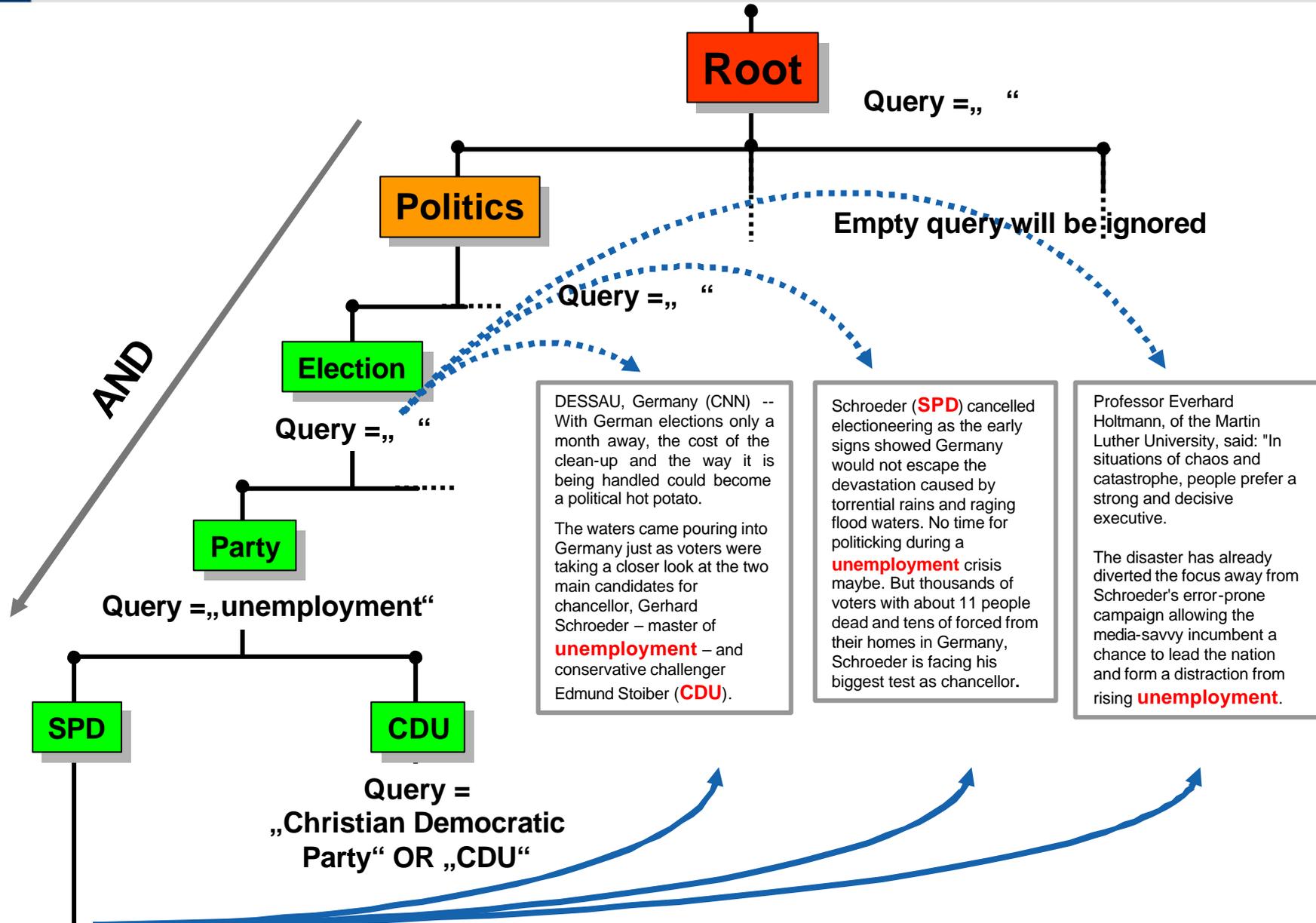
Queries

- ✍ are hierarchical
- ✍ Modification of queries is possible

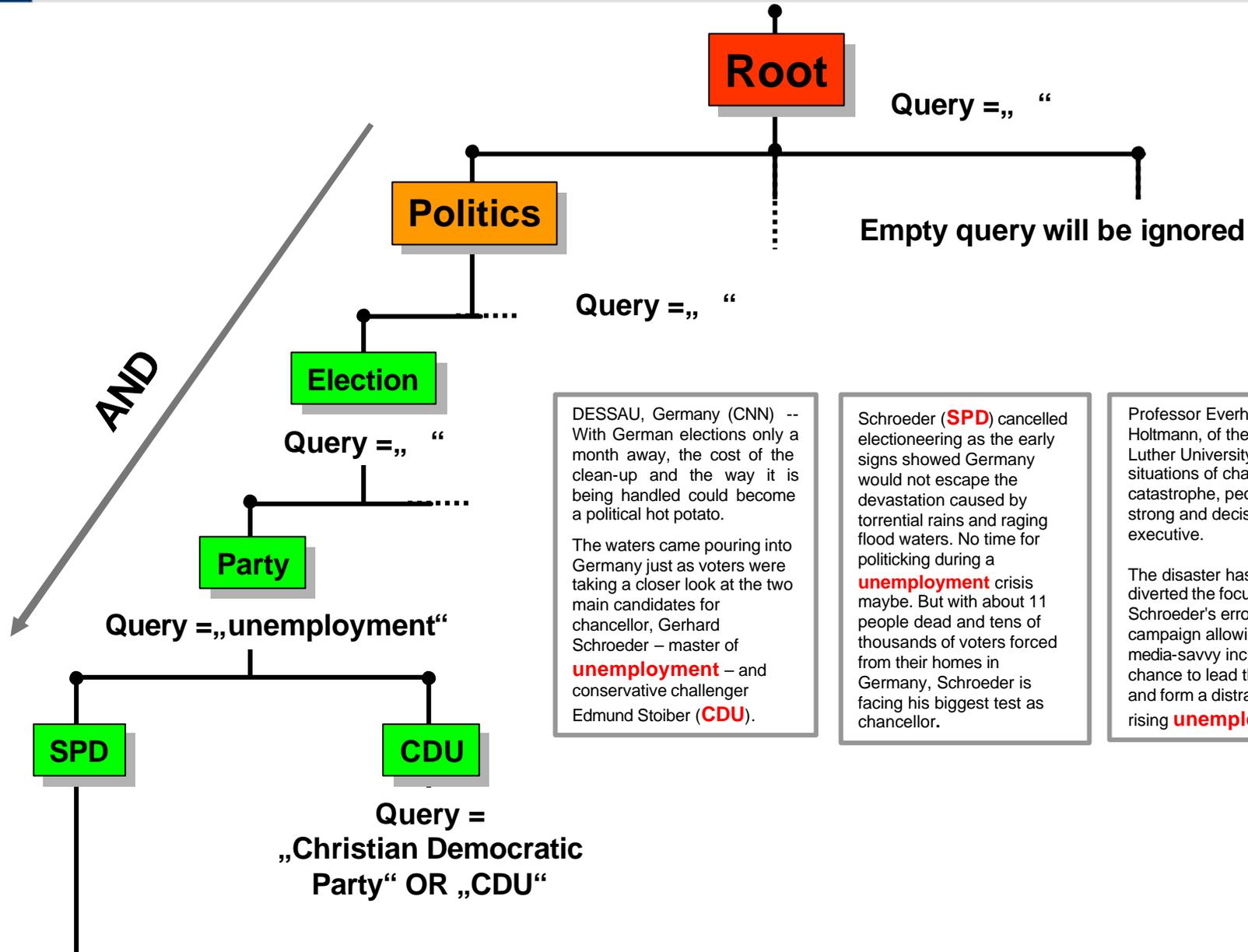
Query based classification :

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

Hierarchical Queries I



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.

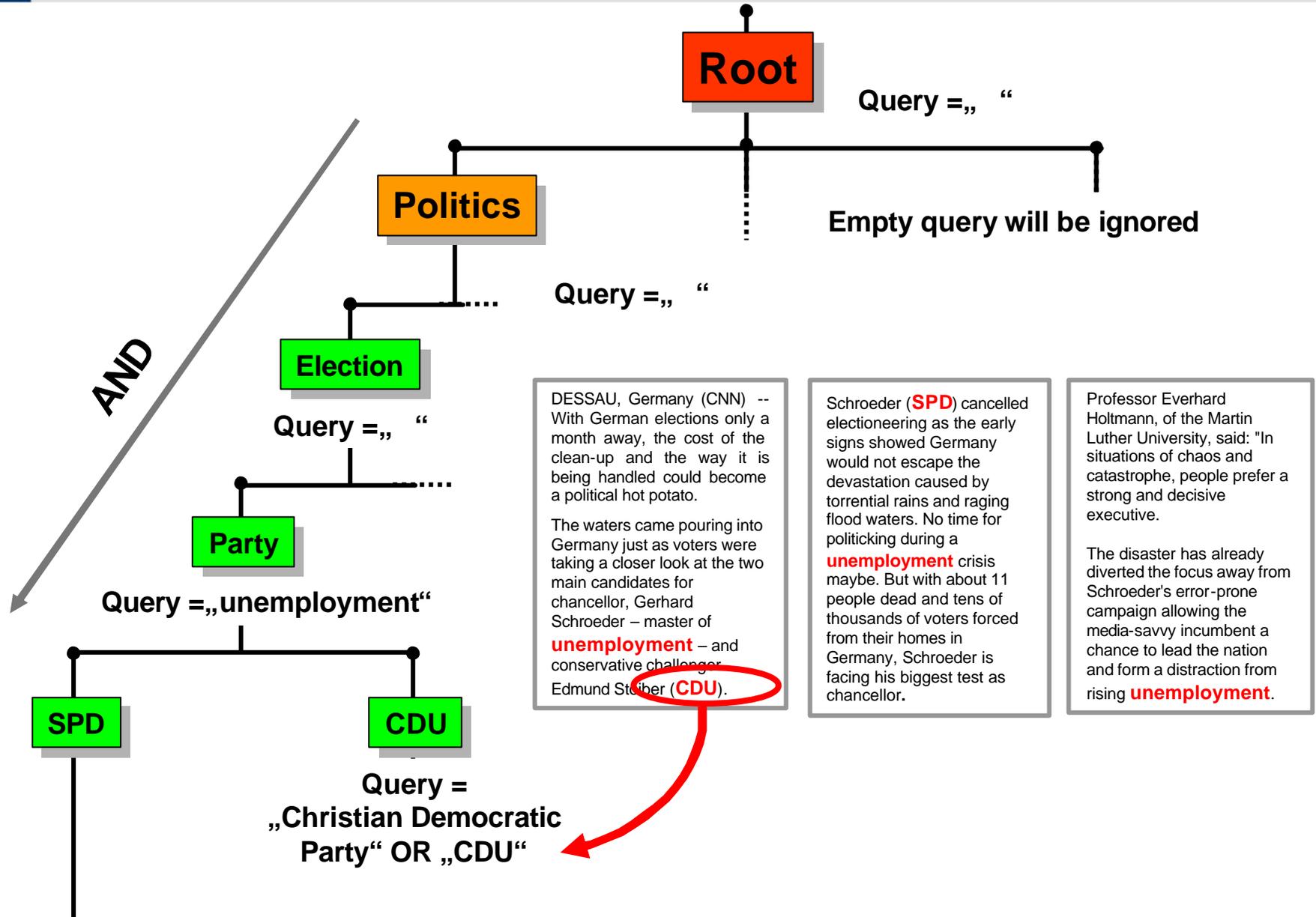
The waters came pouring into Germany just as voters were taking a closer look at the two main candidates for chancellor, Gerhard Schroeder – master of **unemployment** – and conservative challenger Edmund Stoiber (**CDU**).

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.

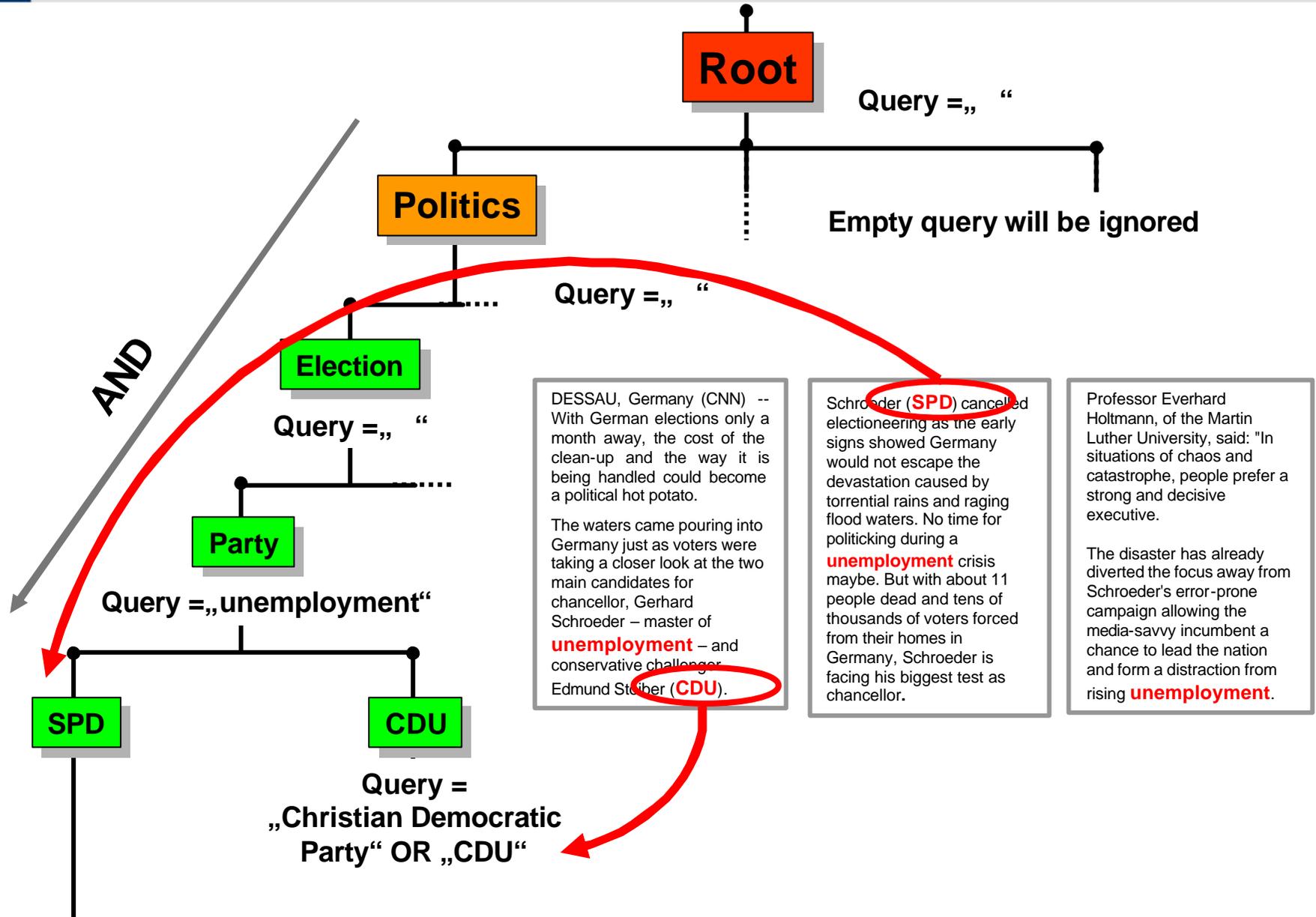
Professor Everhard Holtmann, of the Martin Luther University, said: "In situations of chaos and catastrophe, people prefer a strong and decisive executive.

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**.

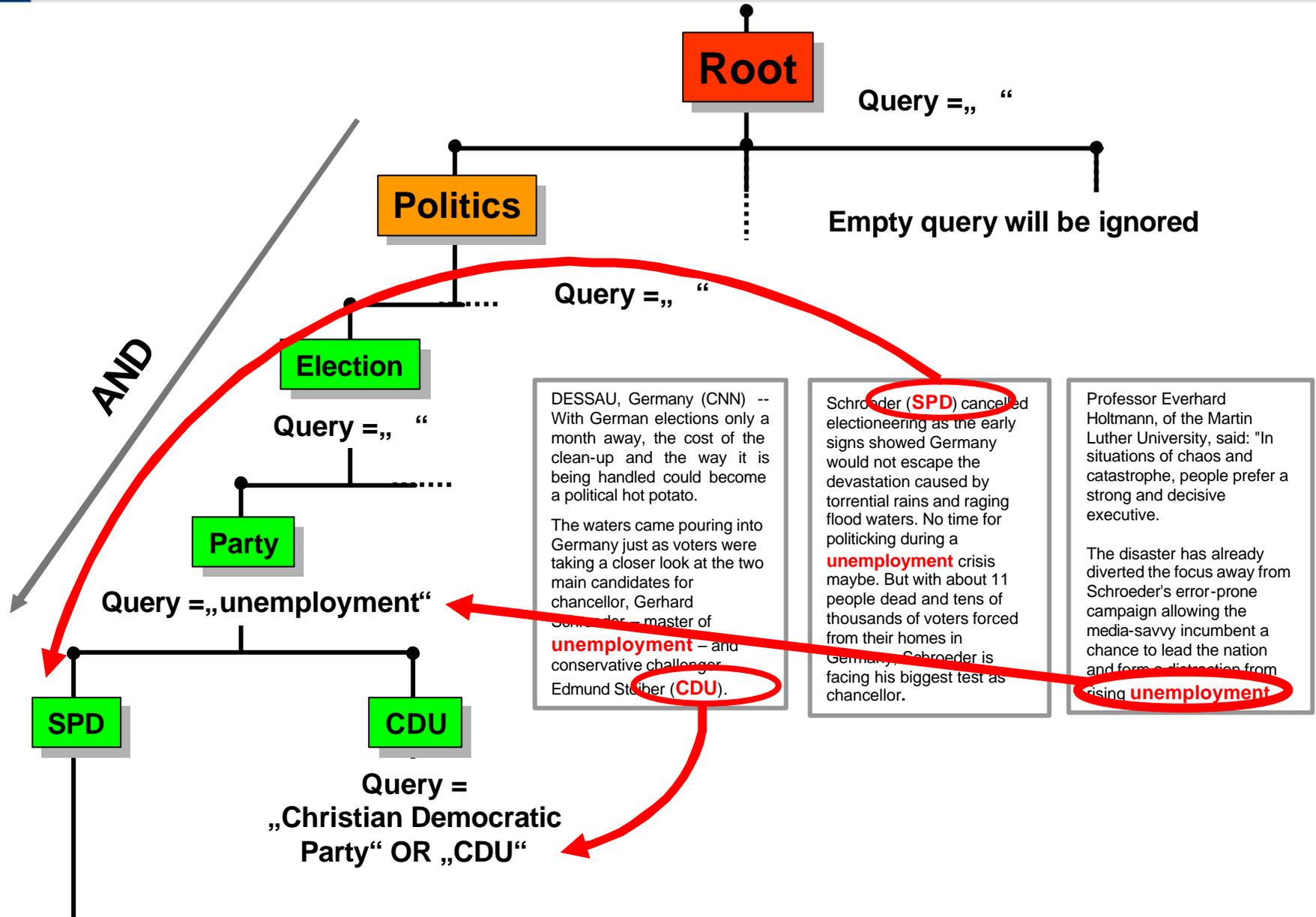
Hierarchical Queries II



Hierarchical Queries II



Hierarchical Queries II



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**

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

Taxonomy Maintenance

SAP Portals Enterprise Portal 5.0 - Microsoft Internet Explorer provided by SAP IT

File Edit View Favorites Tools Help

SAP Welcome Hans-Juergen Richstein Search Personalize: Page | Portal Log Off

My Pages **KM Content** KM Admin KM Monitoring Portal Admin Content Admin Portal Monitoring System Configuration Support

CM Admin Internal Internal **Taxonomy Maintenance** Classification Classification Forms Explorer News News FAQs Inbox Training Editor

Taxonomy Maintenance

You may tune your taxonomy by defining queries to determine the content of taxonomy folders you choose. If the query is complex consider switching from simple view to advanced view.

Taxonomy News

- News
 - Business
 - IBM
 - Big Blue**
 - Microsoft
 - Oracle
 - SAP
 - Politics
 - Sport

Update

Query of News/Business/IBM/Big Blue

Simple View

And Actions

Content Contains Big Blue Delete | Add

Created By Is Richstein, Hans-Jürgen Select... Delete | Add

Or Actions

Title (Displayname) Is *IBM* Delete | Add

Title (Displayname) Is *eBusiness* Delete | Add

Save Check Reset Copy

i Query has been saved for category News/Business/IBM/Big Blue

Applet geladen Local intranet

Classification Inbox for any Taxonomy

The screenshot displays the SAP Enterprise Portal 5.0 interface in Microsoft Internet Explorer. The browser title is "SAP Enterprise Portal 5.0 - Microsoft Internet Explorer provided by SAP IT". The page header includes "Welcome Super Administrator" and navigation links like "Search", "Personalize", "Page", "Portal", and "Log Off". The main navigation bar contains "Portal Admin", "Content Admin", "Portal Monitoring", "System Configuration", "Content Manager", "KM Admin", "KM Monitoring", "Support", and "GFI Democenter". A secondary navigation bar lists "Authoring", "News Maintenance", "Classification Inbox", "Classification Training", "Tax. Query Builder", and "Forms Builder".

The "Classification Inbox" section is active, showing a message: "You may select a taxonomy and view documents not yet classified. You may opt to classify these documents, declassify or reclassify classified documents, or decide never to classify others. If the inbox is large consider further taxonomy training." The "Taxonomy" dropdown is set to "Sailing Equipment".

On the left, a tree view shows the taxonomy structure:

- Documents to Classify
- Never to Classify
- Sailing Equipment
 - Hardware
 - Cleats
 - Fasteners
 - Winches
 - Lines
 - Safety Equipment
 - Sails
 - Genoas
 - Jibs
 - Mainsails
 - Spinnakers
 - Tillers

The "Documents in Sailing Equipment/Sails/Mainsails" table lists the following items:

Documents	Cat.
<input type="checkbox"/> Clear Advantage Mainsail.doc advantage, clear, Doyle, puff, mainsail, Champion, product, week, Line of sails, PHRF	Details 1
<input type="checkbox"/> Dacron Race Mainsails.doc Dacron, RACE MAINS, mainsail, Internal Velcro flap batten, longevity, full top, flap...	Details 1
<input type="checkbox"/> Genesis Mainsail line.doc Sobstad, film, sailmaking, fiber, construction, seam, technology, Adhesive, polyester,...	Details 1
<input type="checkbox"/> J24 Main Sails Overview Basic Overview	Details 5
<input type="checkbox"/> J24 Mainsail J24, Quantum, cross out, mainsail, Dacron, sail design, new sail	Details 1
<input type="checkbox"/> J24 spinnaker sail Good summary of J24 spinnaker	Details 5
<input type="checkbox"/> Mainsail measurement form.htm outhaul car, face of mast, Bearing point, max, top of boom, point of tack fitting	Details 1
<input type="checkbox"/> Mainsail Trim.htm luff tension, mast bend, draft, leech, boom vang, /Sailing/General Information/Sails/images/Mainsail,...	Details 1
<input type="checkbox"/> Mainsail.txt Sails, J24, mainsail, false, jonest, text/plain, text	Details 1
<input type="checkbox"/> Mainsails information.htm full length battens, shelf foot, outhaul, top batten, loose foot, /Sailing/General...	Details 1

At the bottom of the table, it indicates "Page 1/3". Below the table are several action buttons: "<< Classify", "Declassify Here", "Declassify Completely", "Select All", "<< Reclassify", "Never Classify", and "Deselect All".

Table of Contents



Introduction

Example-based Taxonomies

Query-based Taxonomies

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...**

Repositories, Folders, Indices, Taxonomies

