

Query-based Taxonomies

Table of Contents



Introduction

Example-based Taxonomies

Query-based Taxonomies

Planning



Next



Back



Exit



Bookmarks

Multiple Physical Repositories

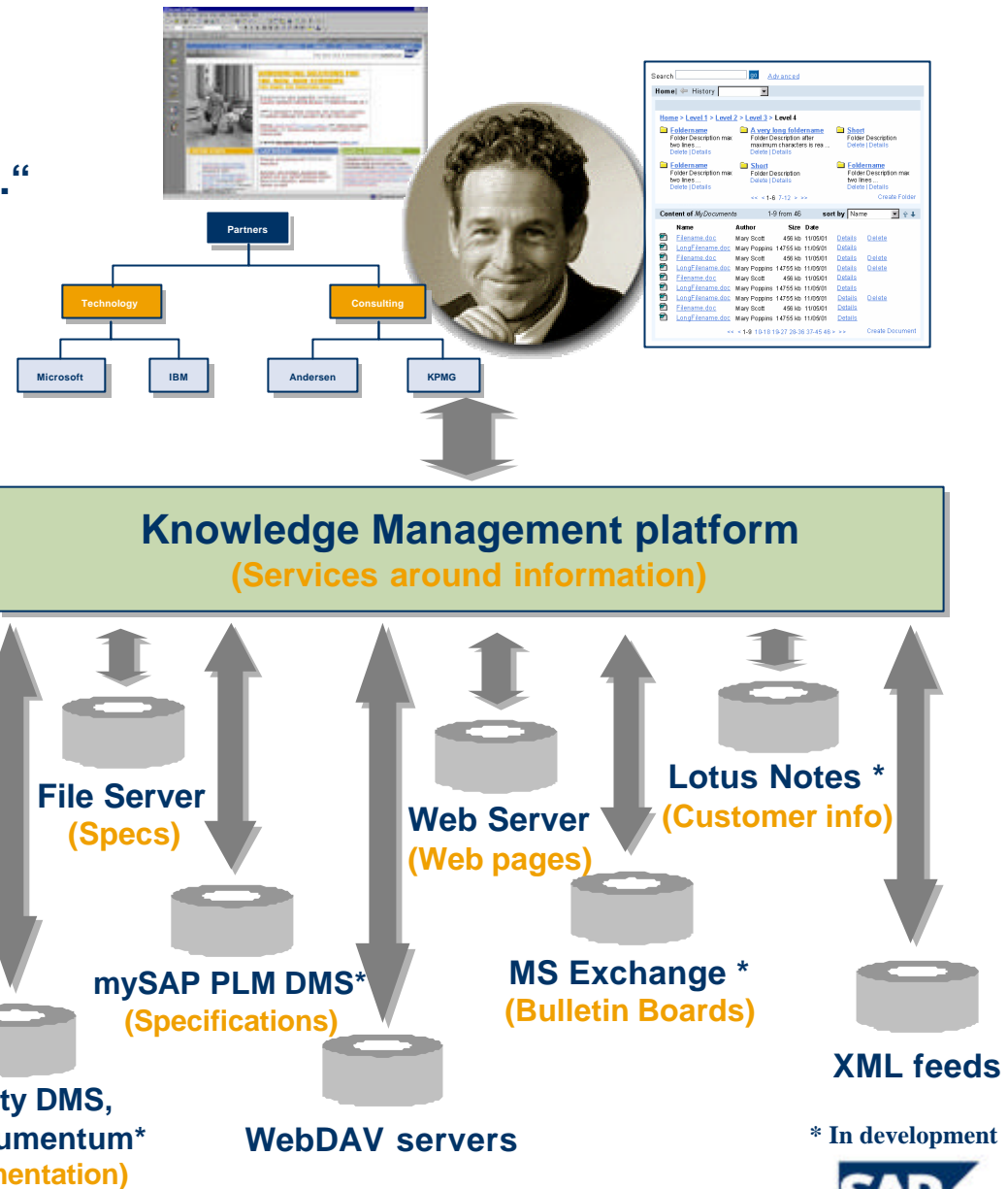
“It is impossible to centralize all the information in a single repository ...”

KM provides

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

KM consists of

- ✍ Content Management
- ✍ Retrieval & Classification
- ✍ Collaboration

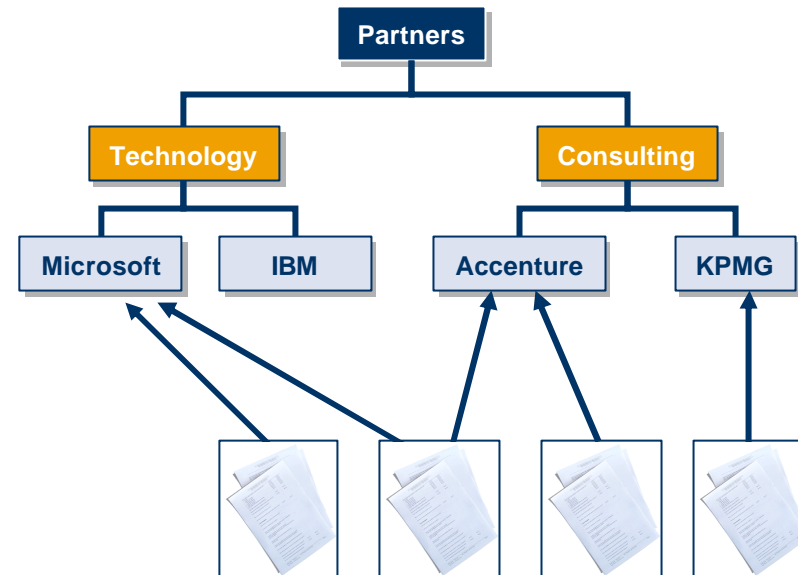


Automatic Document Classification

- ✍ 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

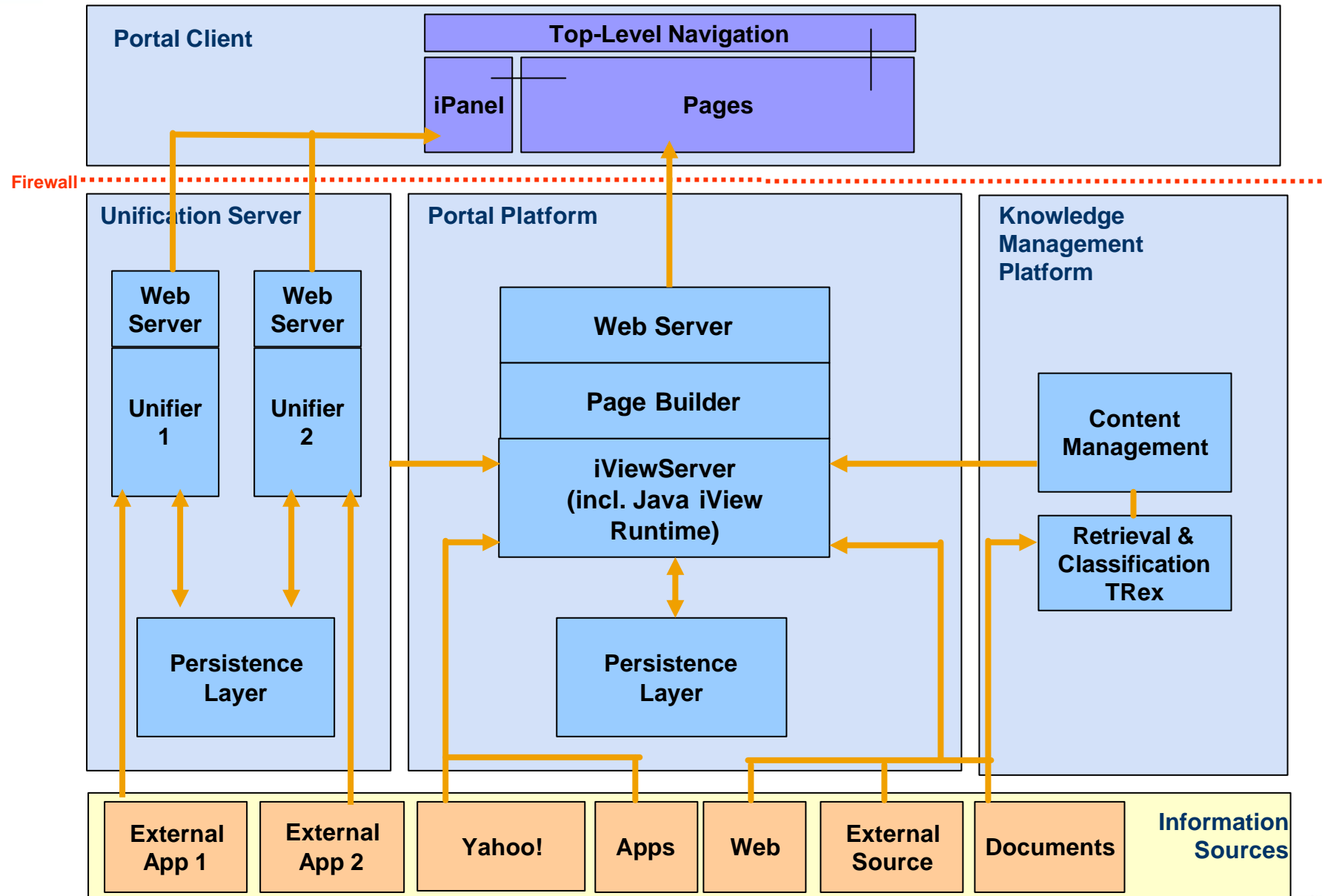


Table of Contents



Introduction

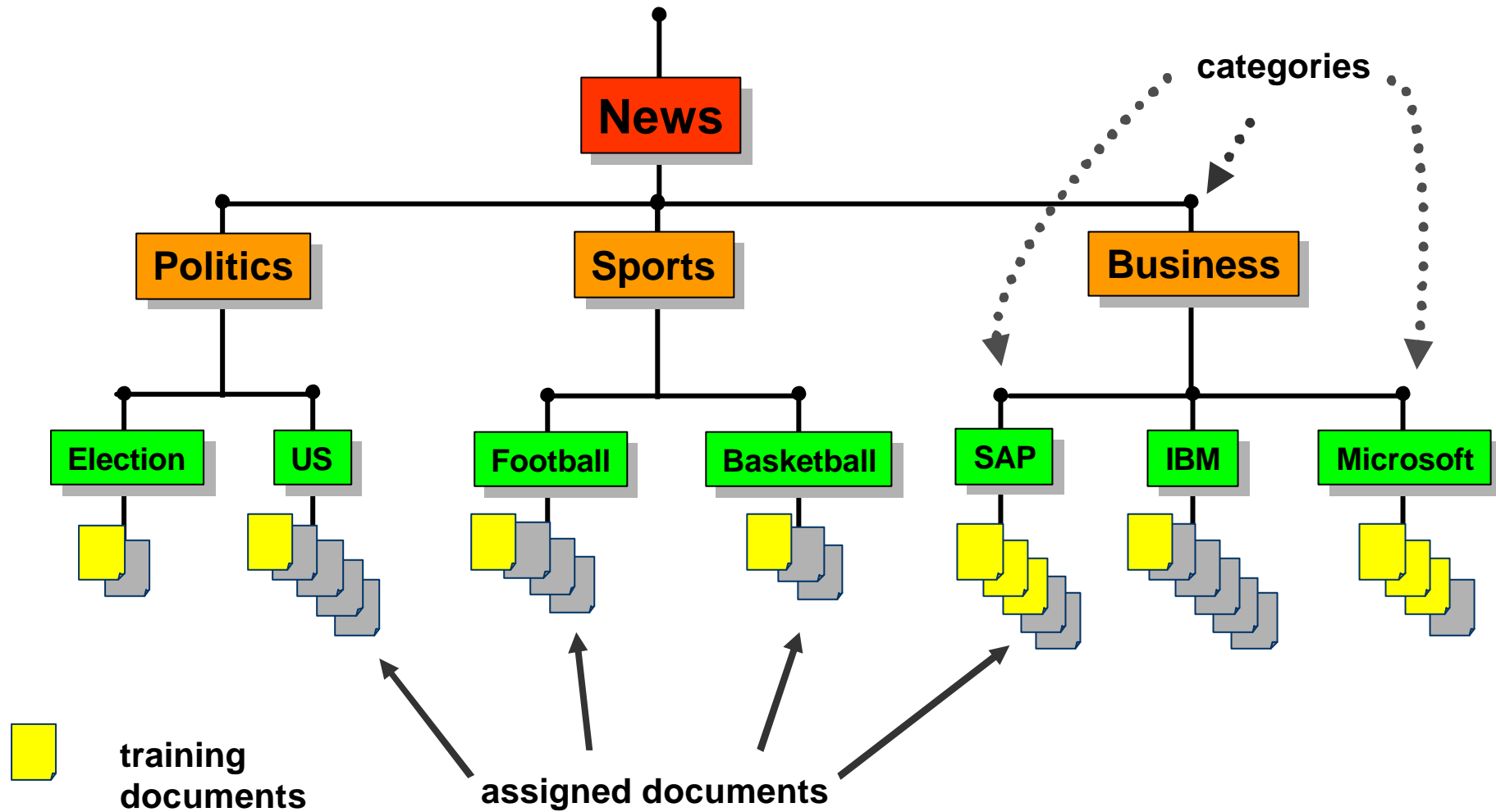
Example-based Taxonomies

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: mytax1

Type: Example-Based

☒ Train with this hierarchy:
test\tax\train Browse...

☒ Include training hierarchy documents in taxonomy

☒ Automatic classification
Category Matching: Medium

OK Cancel

Train EBT with Search Results

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

Welcome Super Administrator

--Collaboration Tools-- Search Personalize: Page | Portal Log Off

Portal Admin Content Admin Portal Monitoring System Configuration **Content Manager** KM Admin KM Monitoring Support GFI Democenter

Authoring News Maintenance Classification Inbox **Classification Training** Tax. Query Builder Forms Builder

Taxonomy Trainer

You may tune your taxonomy by locating documents to train or untrain the taxonomy folder(s) you choose. Refining your search will reduce the amount of manual selection of individual documents.

Choose Taxonomy Folders

root > taxonomies

☐ Kodak ☐ News ☐ Sailing

Locate Documents

cleat Search Show Options

Similar Terms for cleat

☐ CAM ☐ ball ☐ jam ☐ hardware ☐ Aluminum ☐ windage

Check similar term to include in next search or click to search for using one of the terms.

Train Folders Untrain Folders

Search Results for cleat

Sort By Relevance

Select All Deselect All

<input checked="" type="checkbox"/>		Cleats.htm	99%	Last Modified By Kravitz	3.5 KB	12/10/02 2:51:59 PM	See Also Details
				cleat, low profile, windage, Aluminum, hardware, spars, base, steel, relief			
		Sailing > Sailing Equi... > Hardware > Cleats					
<input checked="" type="checkbox"/>		Ball Bearing Cam Cleat.htm	97%	Last Modified By Kravitz	3.5 KB	12/10/02 2:51:59 PM	See Also Details
				CAM, ball, cleat, tooth, pawls, hardware, load applications, engagement, grip, bearings			
		Sailing > Sailing Equi... > Hardware > Cleats					

Table of Contents



Introduction

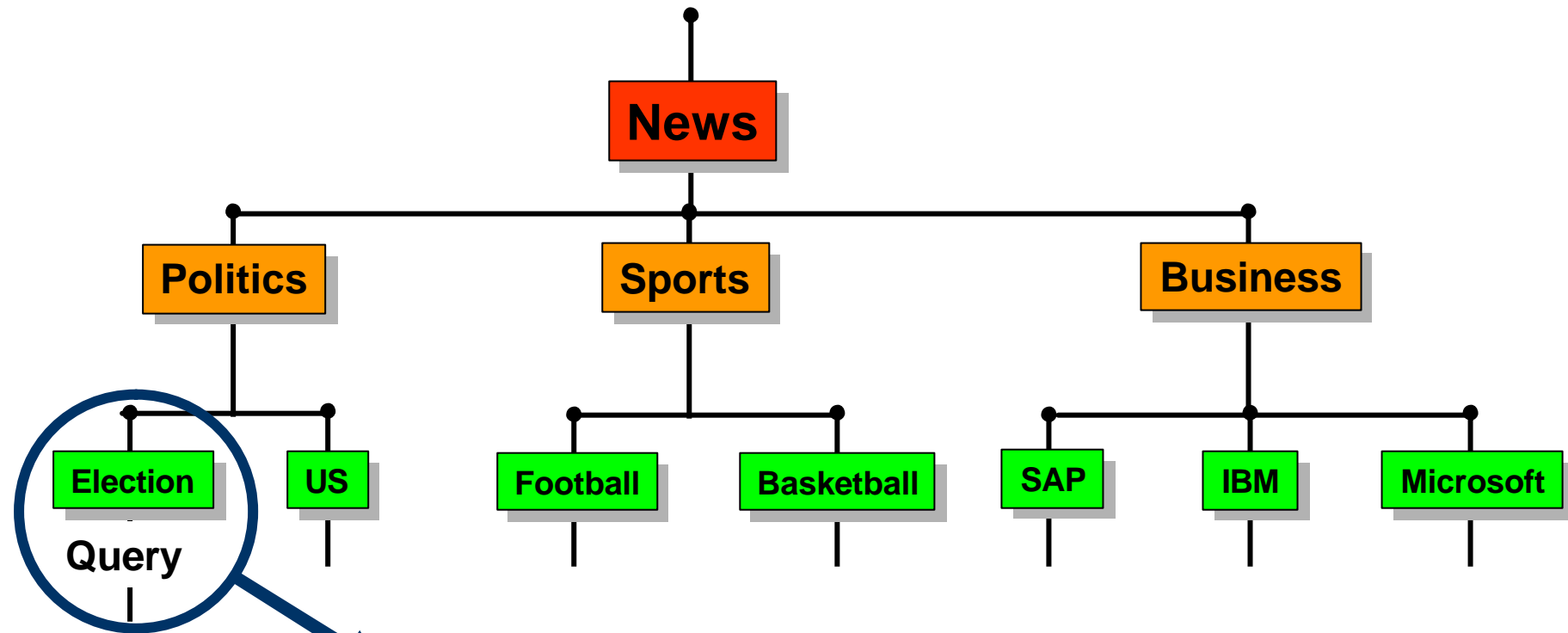
Example-based Taxonomies

Query-based Taxonomies

Planning

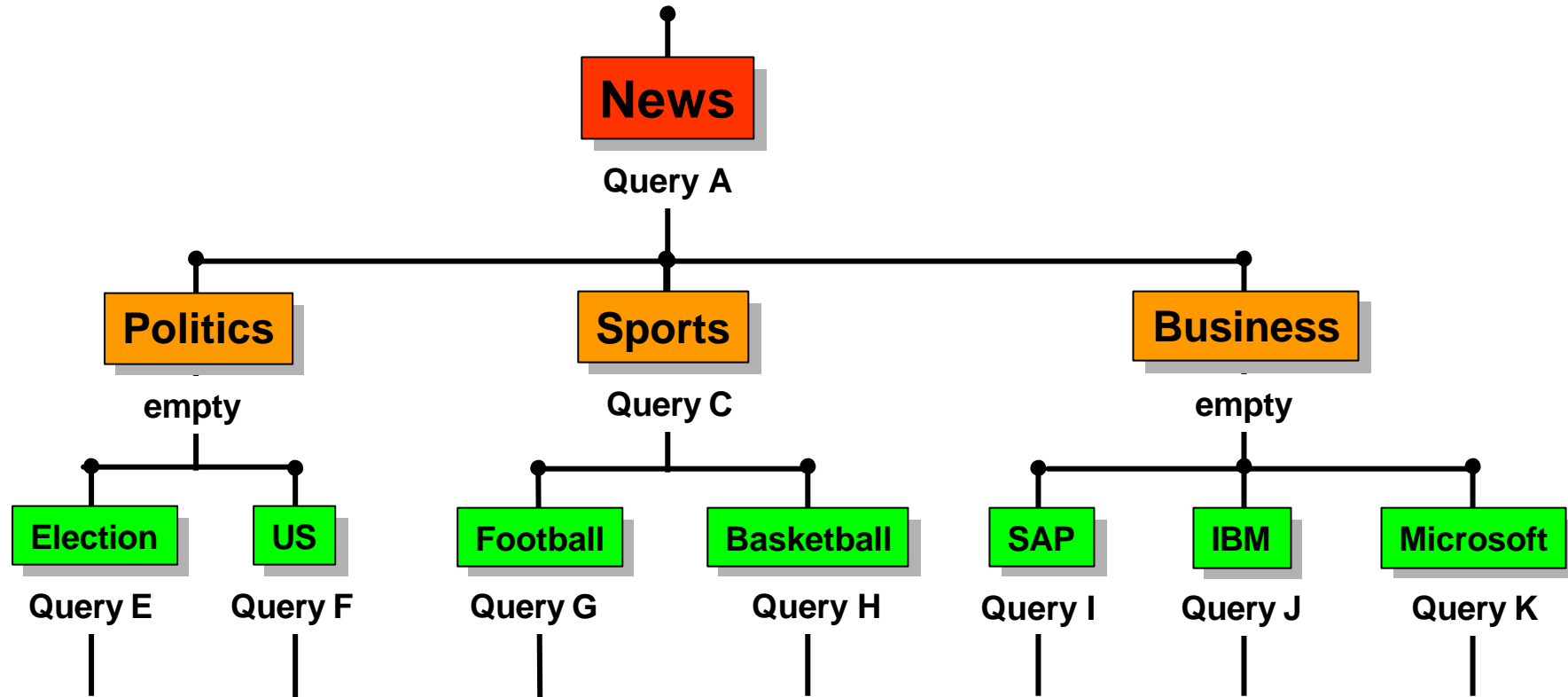
Concept : Query-based taxonomy I

Since SP5 for EP 5.0



Query Election = „unemployment AND party AND strategy“

Concept : Query-based Taxonomy II

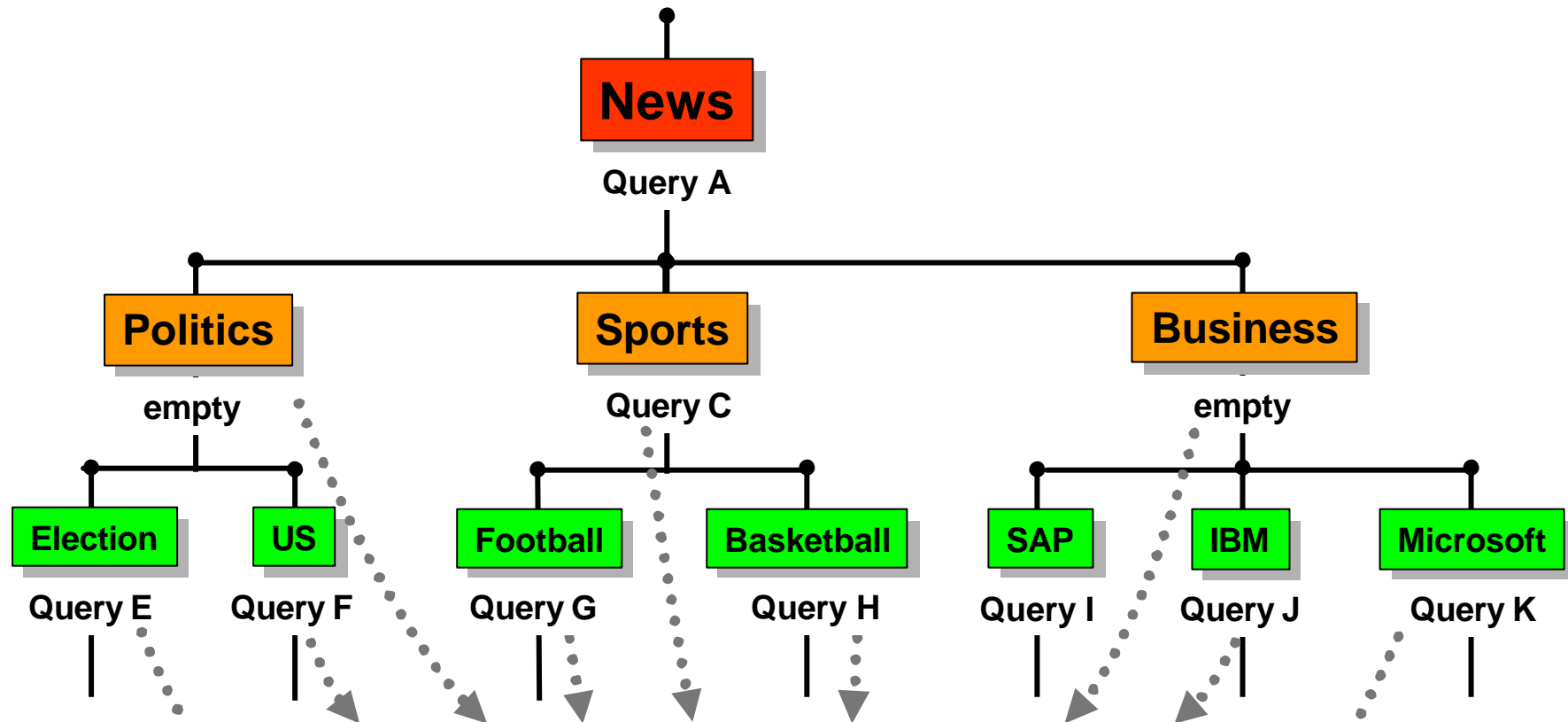


Knowledge is one of the most important production factors. However, the lack of knowledge and information is constantly becoming more obvious. Information exists in numerous file formats that are stored using different kinds of media and in widely distributed storage locations. Because such information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

DOCUMENT

If query matches:
list of classes

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

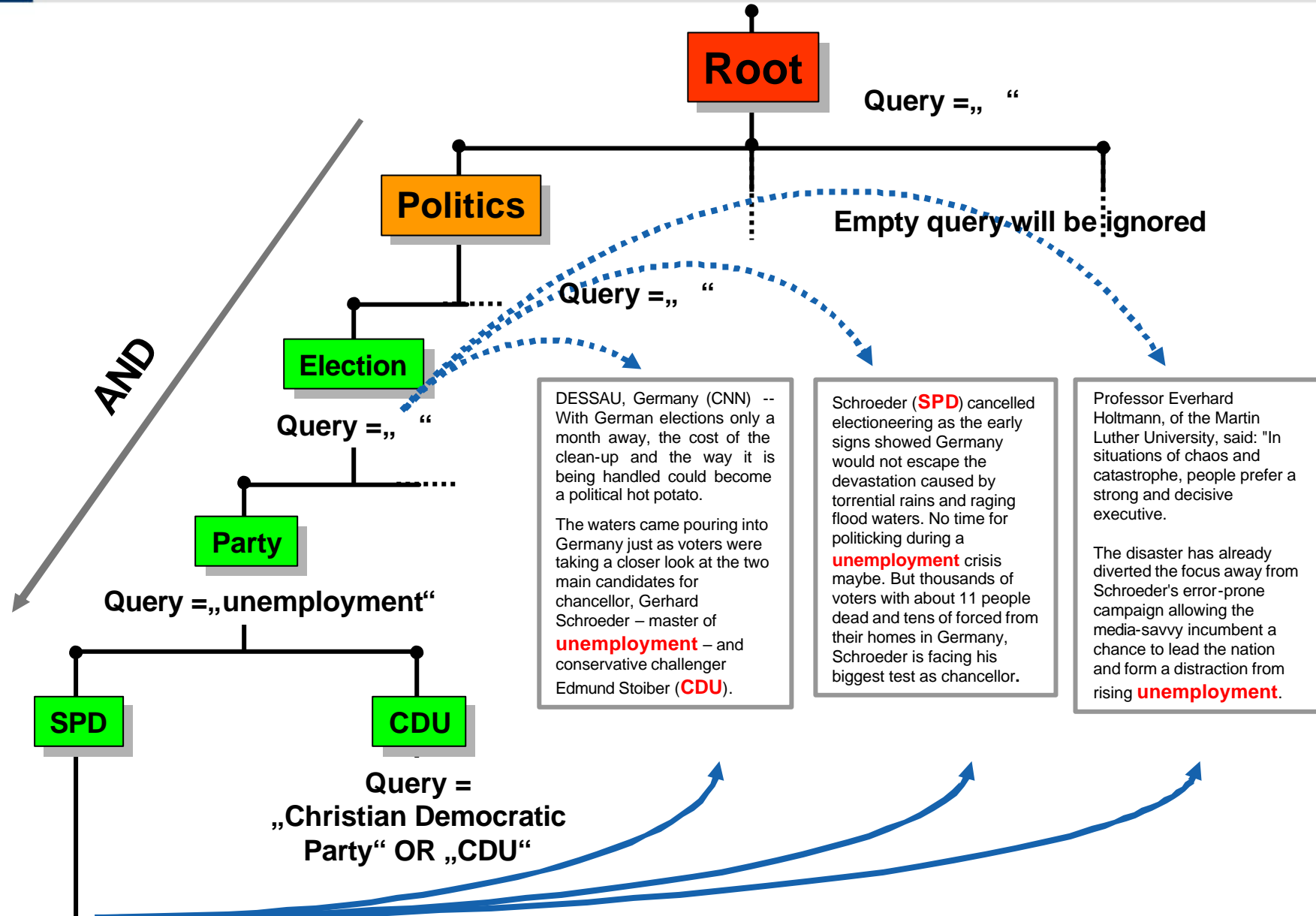
Queries

- ✍ are hierarchical
- ✍ Modification of queries is possible

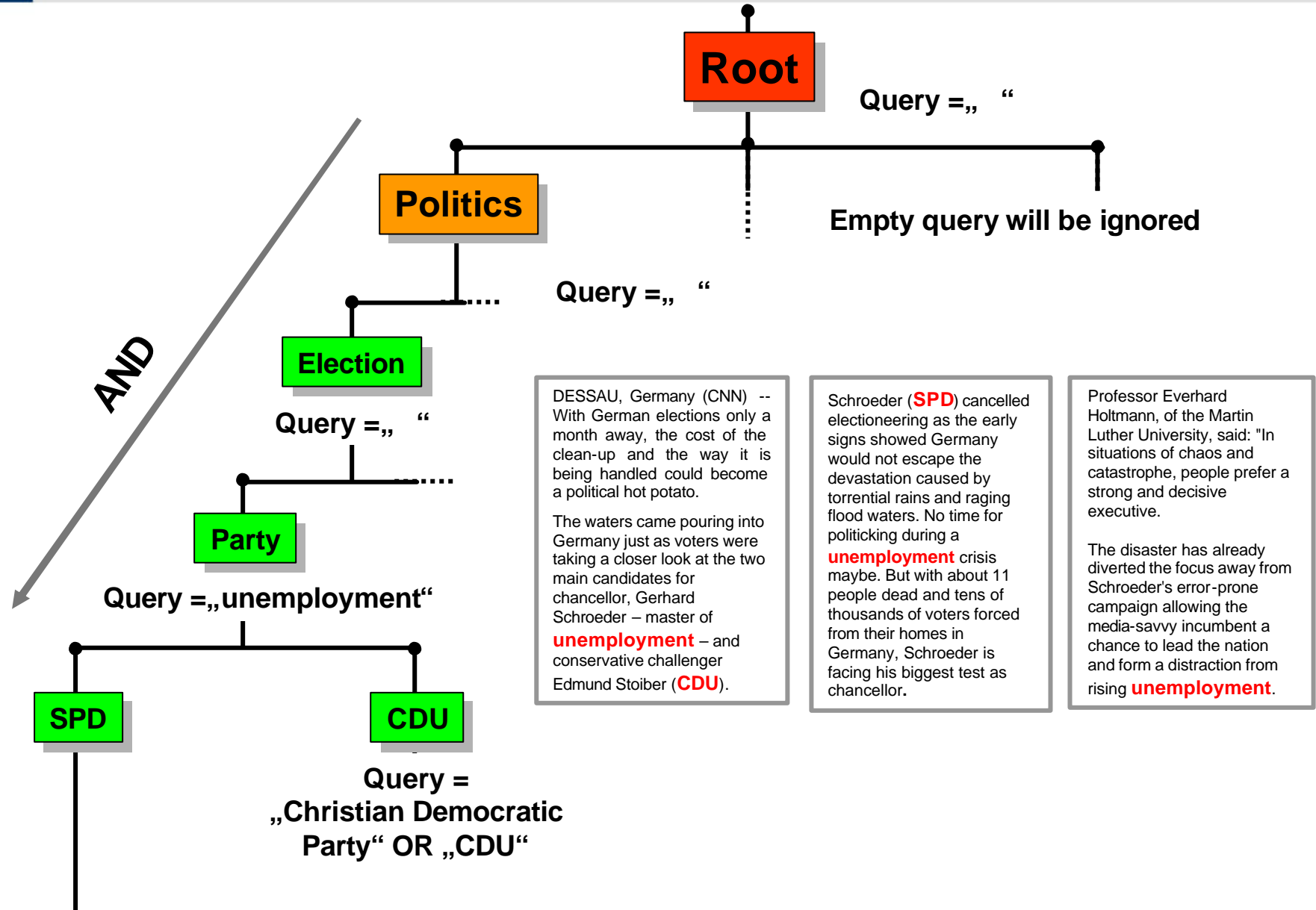
Query based classification :

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

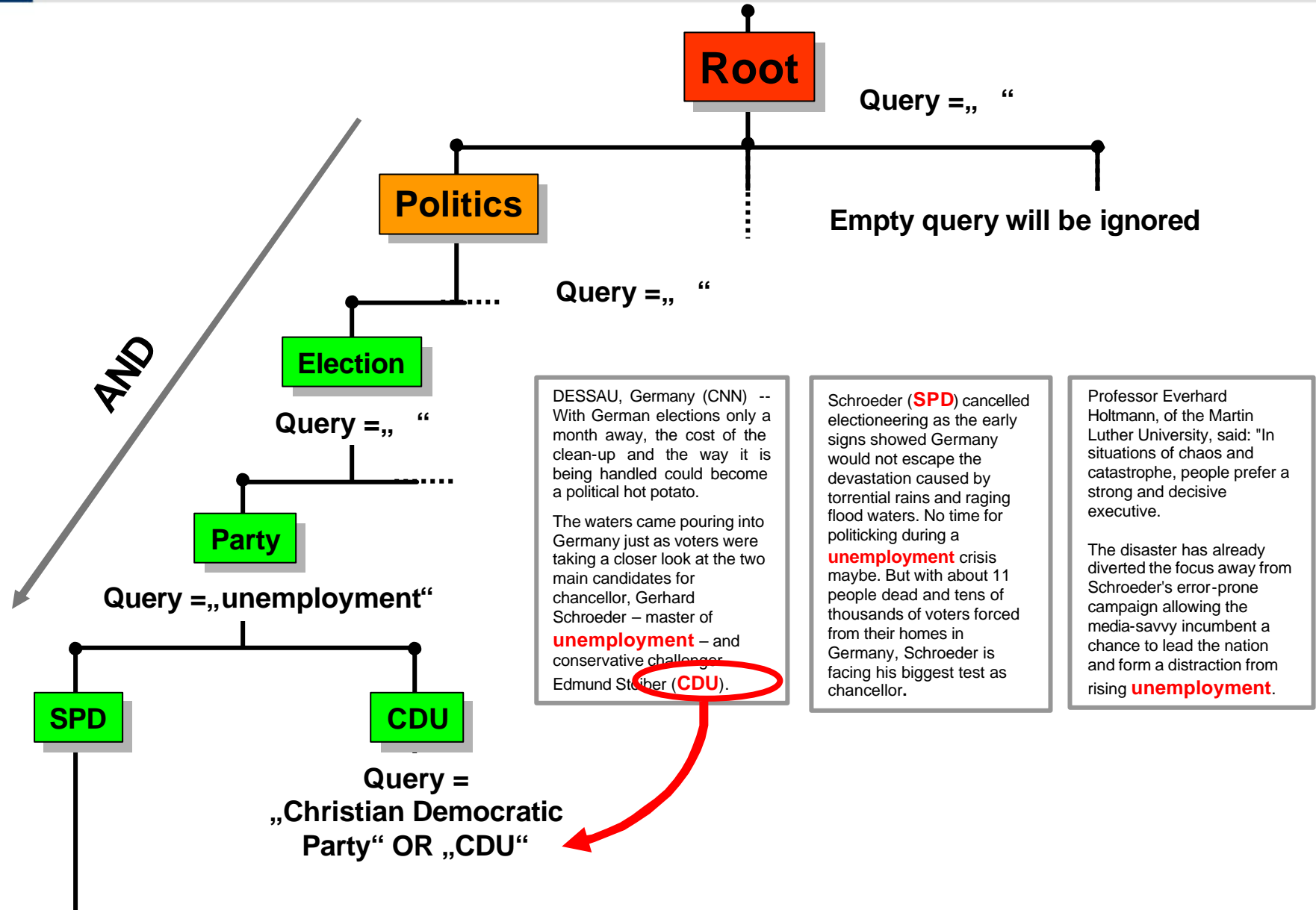
Hierarchical Queries I



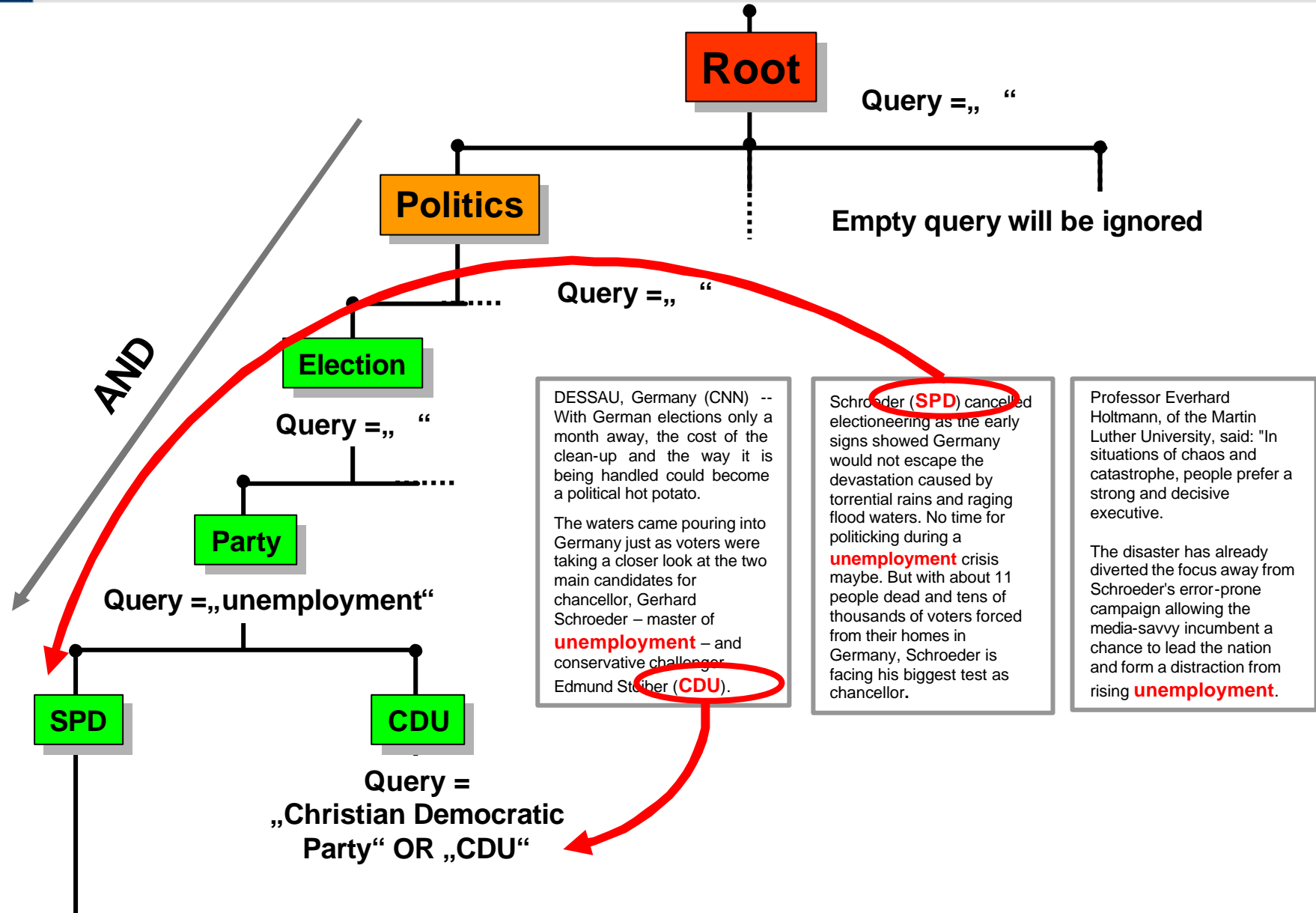
Hierarchical Queries II



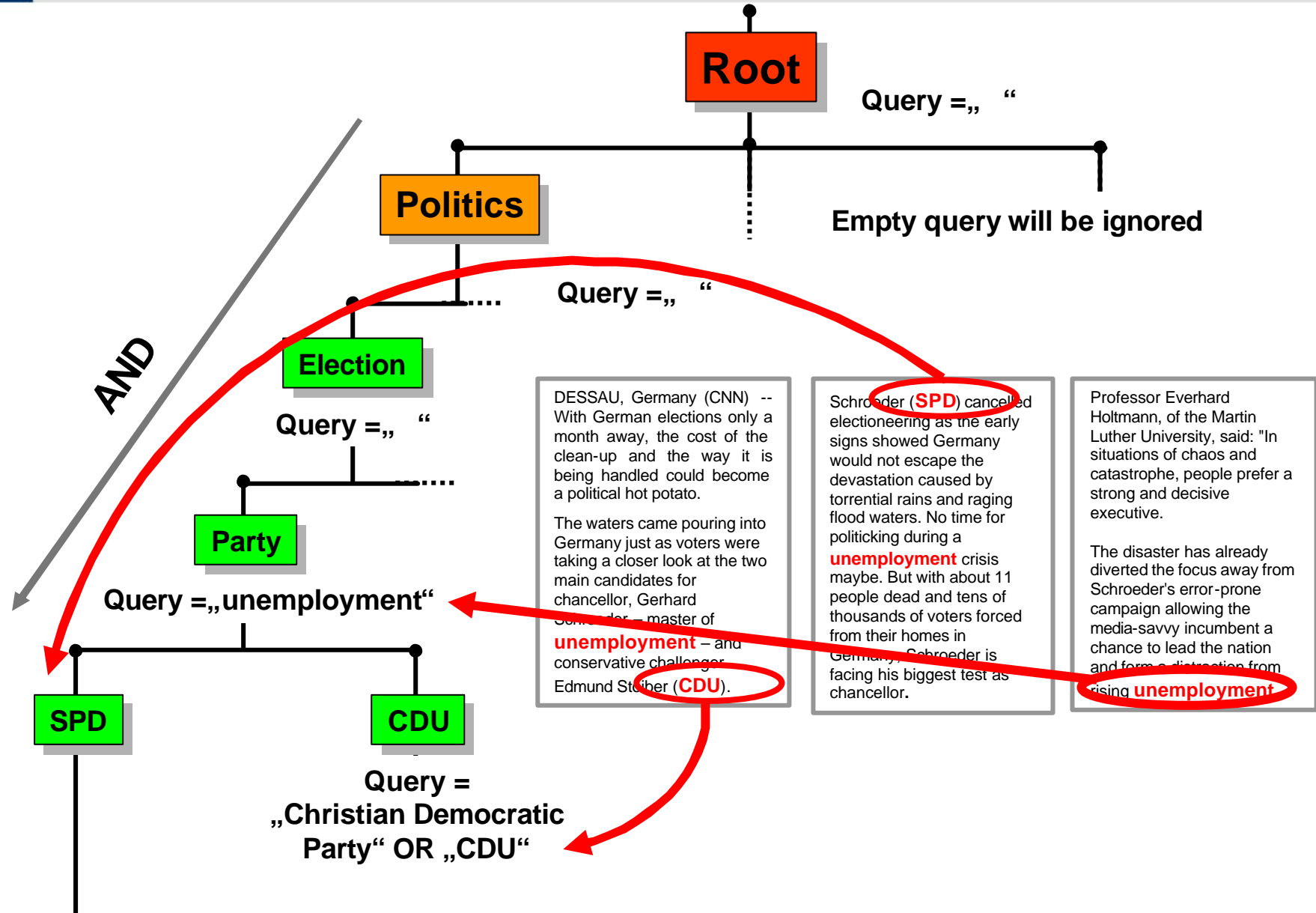
Hierarchical Queries II



Hierarchical Queries II



Hierarchical Queries II



- ✍ **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

The screenshot displays the SAP Portals Enterprise Portal 5.0 interface within a Microsoft Internet Explorer browser. The title bar reads "SAP Portals Enterprise Portal 5.0 - Microsoft Internet Explorer provided by SAP IT". The browser's address bar is empty, and the menu bar includes File, Edit, View, Favorites, Tools, and Help. The SAP logo is in the top left corner, and a welcome message "Welcome Hans-Juergen Richstein" is in the top right. A search bar and links for "Personalize: Page | Portal | Log Off" are also present. A navigation bar contains various menu items: My Pages, KM Content, KM Admin, KM Monitoring, Portal Admin, Content Admin, Portal Monitoring, System Configuration, and Support. Below this, a secondary bar highlights "Taxonomy Maintenance" among other options like CM Admin Explorer, Internal News, Internal FAQs, Classification Inbox, Classification Training, and Forms Editor. The main content area is titled "Taxonomy Maintenance" and includes a brief instruction: "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." On the left, a tree view shows the hierarchy: News > Business > IBM > Big Blue. The main panel, titled "Query of News/Business/IBM/Big Blue", shows a "Simple View" configuration. It features two rows of criteria: the first row has "Content" containing "Big Blue", and the second row has "Created By" equal to "Richstein, Hans-Jürgen". Below these, an "Or" section adds two more criteria: "Title (Displayname)" is "*IBM*" and "Title (Displayname)" is "*eBusiness*". At the bottom of the query panel are buttons for "Save", "Check", "Reset", and "Copy". A status message at the bottom left states: "Query has been saved for category News/Business/IBM/Big Blue". The browser's status bar at the bottom shows "Applet geladen" and "Local intranet".

Classification Inbox for any Taxonomy

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

Welcome Super Administrator

Portal Admin Content Admin Portal Monitoring System Configuration **Content Manager** KM Admin KM Monitoring Support GFI Democenter

Authoring News Maintenance **Classification Inbox** Classification Training Tax. Query Builder Forms Builder

Classification Inbox

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.

Taxonomy Sailing Equipment

☐ Documents to Classify
☐ Never to Classify

☐ Sailing Equipment

- ☐ Hardware
 - ☐ Cleats
 - ☐ Fasteners
 - ☐ Winches
- ☐ Lines
- ☐ Safety Equipment
- ☐ Sails
 - ☐ Genoa's
 - ☐ Jibs
 - ☐ Mainsails
 - ☐ Spinnakers
- ☐ Tillers

Clear Selection

Documents in Sailing Equipment/Sails/Mainsails

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

Page 1/3

<< Classify | Declassify Here Declassify Completely Select All
 << Reclassify Never Classify Deselect All

Info

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

