



Fuzzy Search – Identification of anything, anytime, anywhere

Dr. Mario Lenz, Chief Technology Officer
Berlin, February 13th, 2007



- About empolis
- The empolis Information Access Suite
- Fuzzy search for persons as an example
- Discussion

empolis is part of Bertelsmann AG



Bertelsmann AG

88,500 employees (FY 2005)

\$23 billion revenue (FY 2005)



- Information Logistics applies the just-in-time philosophy known from the production world to the commodity “information”:
 - Information and communication processes become efficient, controllable and geared to the needs of the user.
 - Value-creating processes receive maximum support.

Right Information
Right Time
Right Person



Information Logistics made by empolis



empolis

SIEMENS



Wolters Kluwer
Belgium

Schlumberger

FUJITSU COMPUTERS
SIEMENS



International
Accounting Standards
Board®

T · Mobil ·



BOSCH



AIRBUS

ARCOR



SATURN

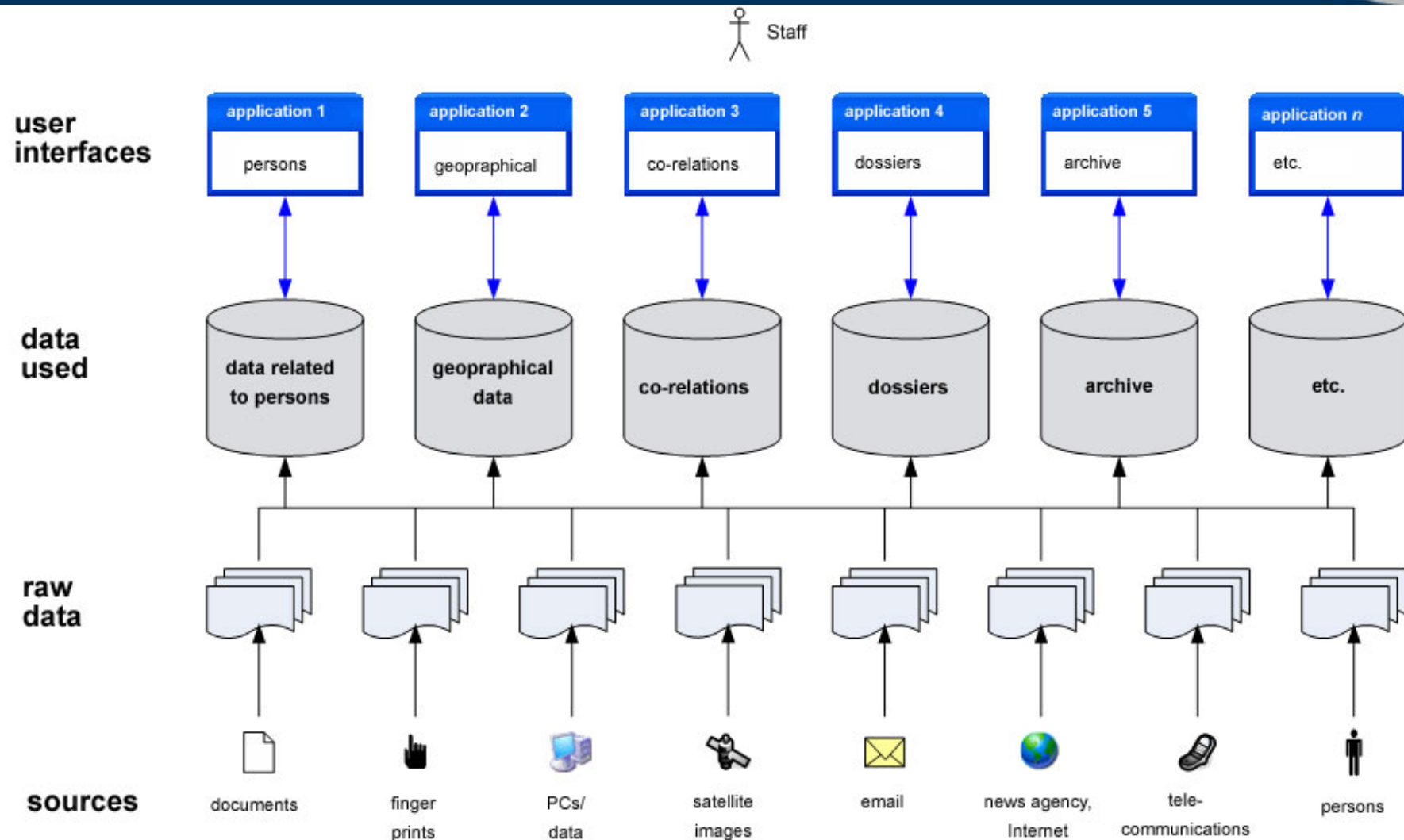
AVON

Media Markt

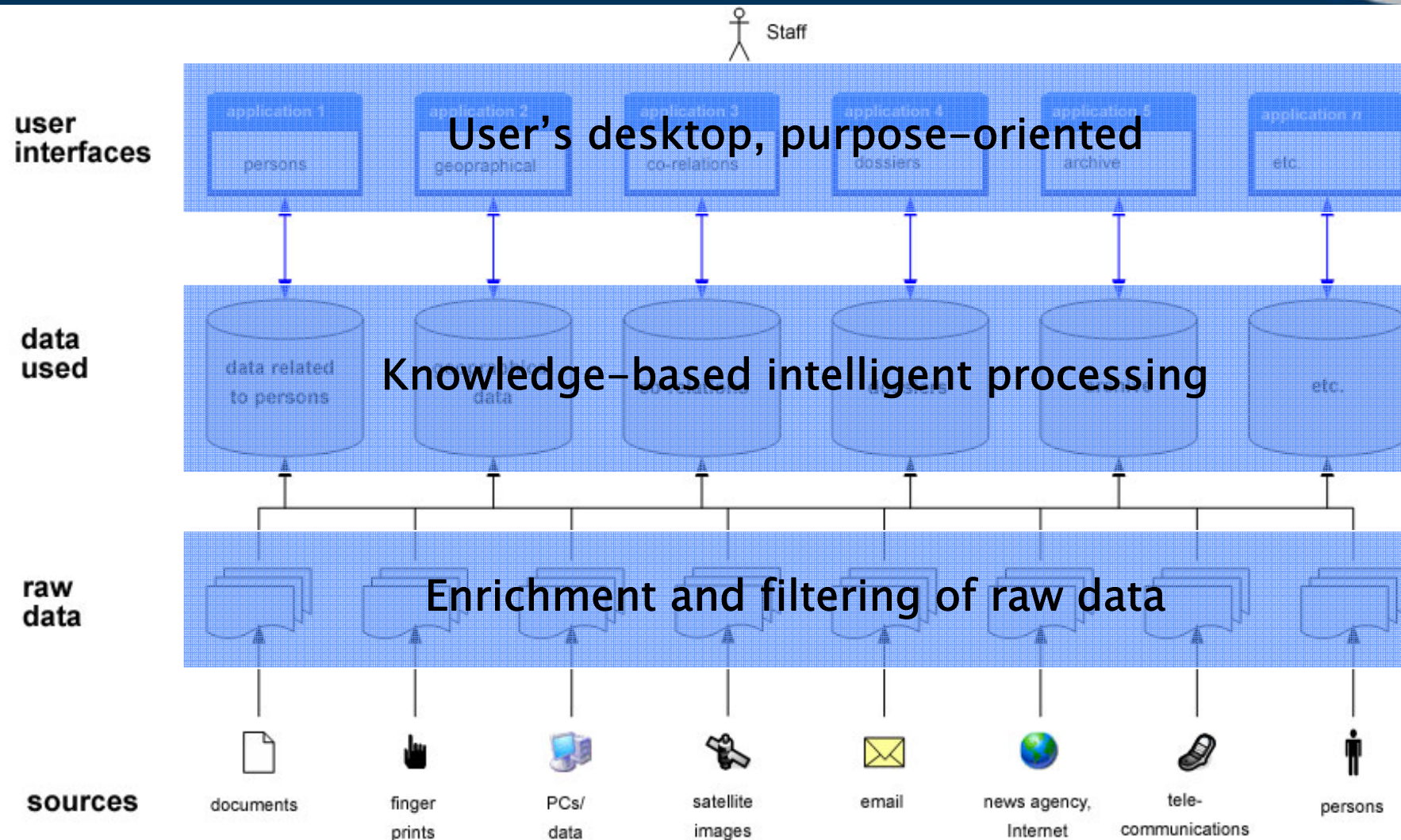
Weltbild.de

OTTO

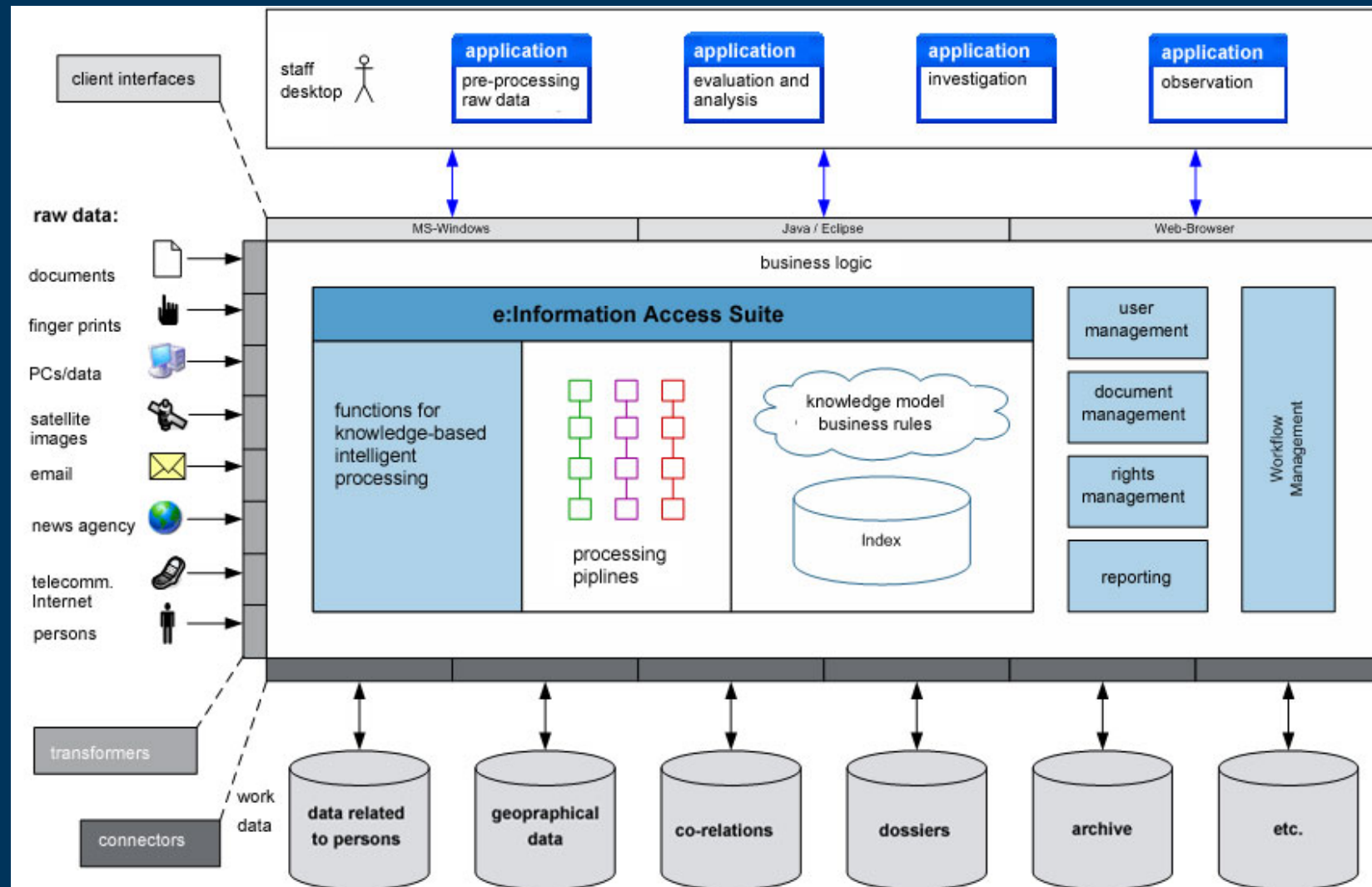
Security Information Systems Today



Tomorrow's Vision



e:IAS – Architecture

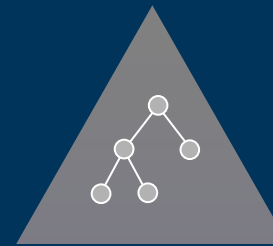


Example: A Simple (?) Search

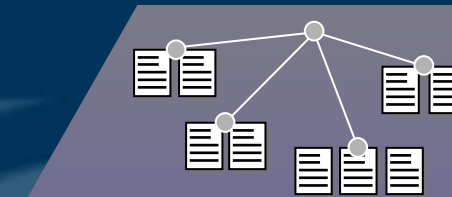


Broad Range of Techniques Available

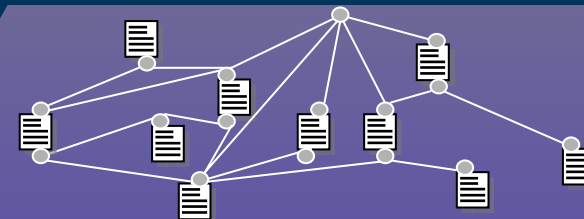
Decision trees (scripts) provided guided procedures through defined processes and workflows based on search technology



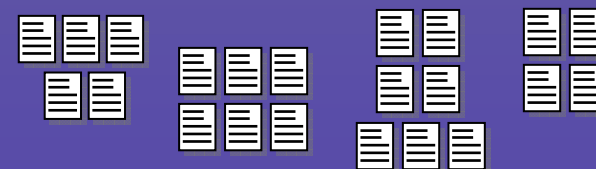
Semantic search and classification are based on ontological knowledge to utilize relationships and similarities of concepts



Associative search is based on sound mathematical models to identify relationships among concepts by statistical methods



Full text search operates on the level of formal boolean logic to identify matching documents



Status Quo – Selected Features

Topic Maps, Ontologies,
Semantic Maps

Dialog-oriented search

Rule-based processing

Case-based reasoning,
Decision trees

Related terms

Summarization,
Semantic fingerprint

Similar documents

Query Manager

Fully automatic classification

Multi-dimensional
organizing principles

Fulltext
Search

Structured
Search

Associative
Search

User agents and profiles

Information Extraction

Content Pre-processor

Clustering

Multi-dimensional feature-based
match making

Term Highlighting

Search for documents, persons,
profiles, experts, news etc.

Personalized news channels

Semantic search for content

Sort by relevance and/or user's
interest

Fault-tolerant search

Automated information
networking

Automatic distribution of
information

320d, EZ 04/05, 23tkm, 150 PS, black/metallic, Xeon, Navi, Leder, VHB 19,5t€



Brand	BMW
Model	320d
Fuel Type	Diesel
Power	110 KW
Color	Black
Equipment	Xenon, Navi, Leather
Built	2005
Mileage	23.000 km
Price	19.500 €

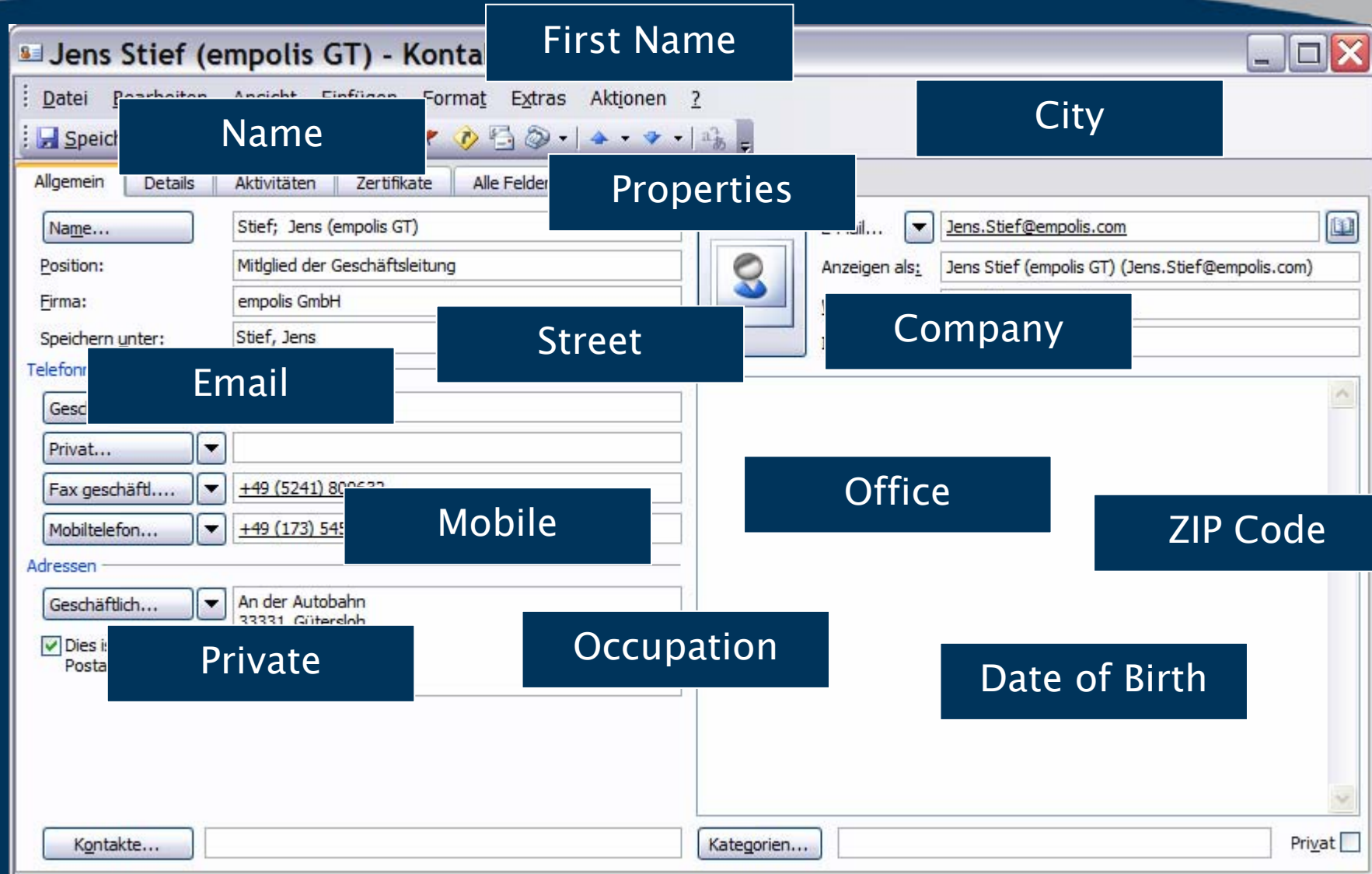
Match-Making Feature by Feature

	Query		Proposal
Brand	BMW		Mercedes
Model	320d		CDI 220
Fuel Type	Diesel		Diesel
Power	110 KW		105 KW
Color	Black		Silver
Equipment	Xenon, Navi, Leather		Navi, AirCon.
Built	2005		2006
Mileage	23.000 km		15.000km
Price	19.500 €		22.000 €

Match-Making Feature by Feature

	Query		Proposal
Brand	BMW	70%	Mercedes
Model	320d	80%	CDI 220
Fuel Type	Diesel	100%	Diesel
Power	110 KW	95%	105 KW
Color	Black	30%	Silver
Equipment	Xenon, Navi, Leather	33%	Navi, AirCon.
Built	2005	100%	2006
Mileage	23.000 km	100%	15.000km
Price	19.500 €	80%	22.000 €
Total		76%	

Feature-based Match-Making to Identify Persons

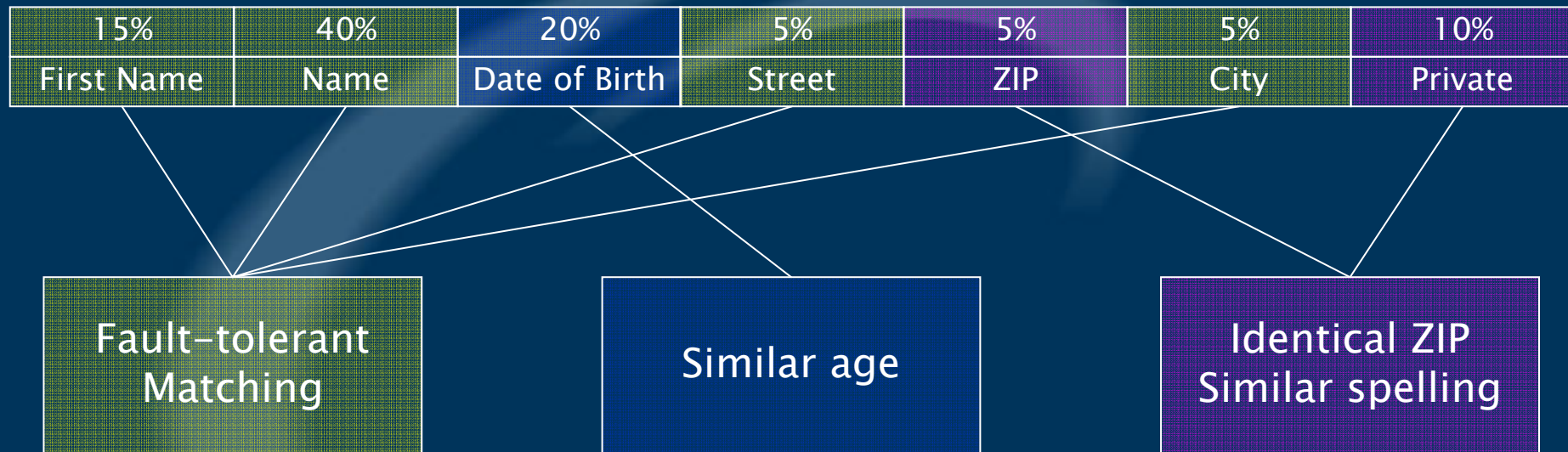


The screenshot shows a contact form for 'Jens Stief (empolis GT)'. The form is divided into several sections, each with a label in a dark blue box. The labels are: 'First Name' (pointing to 'Jens'), 'City' (pointing to 'Güterloh'), 'Name' (pointing to the full name 'Stief, Jens (empolis GT)'), 'Properties' (pointing to the 'Mitglied der Geschäftsleitung' position), 'Street' (pointing to 'An der Autobahn 33331 Güterloh'), 'Company' (pointing to 'empolis GmbH'), 'Email' (pointing to 'Jens.Stief@empolis.com'), 'Office' (pointing to the 'Geschäftlich...' address type), 'Mobile' (pointing to the 'Mobiltelefon...' field), 'ZIP Code' (pointing to '33331'), 'Private' (pointing to the 'Privat...' address type), 'Occupation' (pointing to the 'Mitglied der Geschäftsleitung' position), and 'Date of Birth' (pointing to the 'Geburtsdatum...' field). The form also includes a 'Speichern unter:' field with the value 'Stief, Jens', a 'Kontakte...' button, and a 'Kategorien...' button.

Field	Value
Name	Stief, Jens (empolis GT)
First Name	Jens
City	Güterloh
Properties	Mitglied der Geschäftsleitung
Street	An der Autobahn 33331 Güterloh
Company	empolis GmbH
Email	Jens.Stief@empolis.com
Office	Geschäftlich...
Mobile	+49 (173) 549...
ZIP Code	33331
Private	Privat...
Occupation	Mitglied der Geschäftsleitung
Date of Birth	Geburtsdatum...

Match-Making by Dynamically Weighting Features

- Case-Based Reasoning technology for match-making
 - Multi-dimensional comparison
 - Different types of features (String, Range, numeric etc.)
 - Different measures on feature level
 - Dynamic weighting of features



Example: "Schmidt"

Kontaktperson

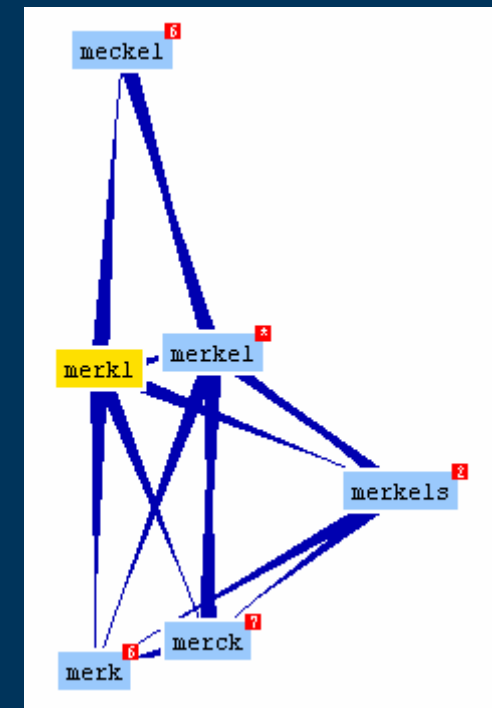
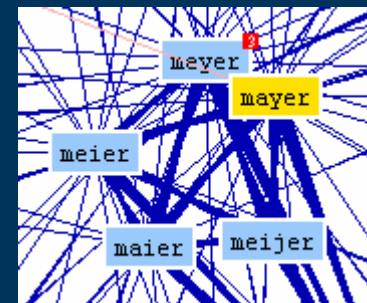
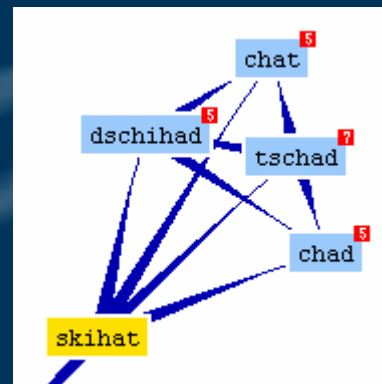
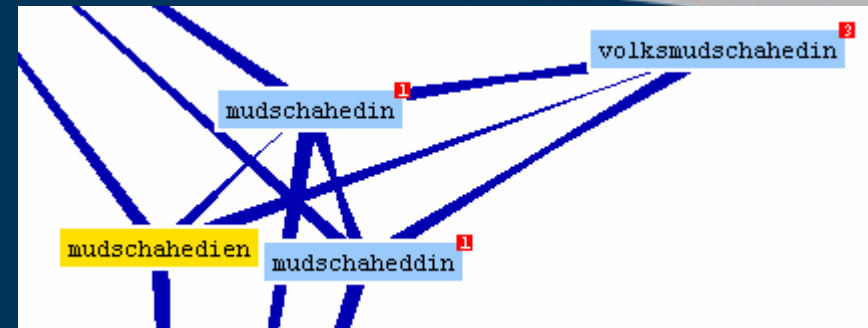
normales Gewicht

Attribut	Anfrage	1. Treffer	100%	2. Treffer	50%	3. Treffer	50%
Kontaktperson	schmidt	Herr Schmidt - ML		Frau Schmitt		Herr Andreas Magnus Schmid	
Kontaktperson (normalisiert)	schmidt	schmidt	100%	schmitt	0%	schmid	0%
Kontaktperson (soundex)	S86	S86	100%	S86	100%	S86	100%

Match Making for Names

Different technologies and options

- Similar spelling (typos)
- Similar sound (soundex)
- Similar transcription (e.g. for Arabic)
- Framework to easily integrate specialized applications and methods
 - E.g. “Kölner Phonetik”





Questions?

Mario.Lenz@empolis.com

**Thank you for your attention !
Visit us on Level B, Booth 19/20**