

CONVERA™

CONVERA OVERVIEW



IAEA

Trieste - November 2004



CONVERA™

● ● ● Agenda

- [About CONVERA](#)
- Introduction
- CONVERA Platform Infrastructure
- CONVERA's Retrieval Technologies
- Visualisation
- Automatic Categorization Technologies
- Dynamic Classification Technologies
- What we do / What's Third Party
- Product demonstration



Convera Confidential

Convera Corporation

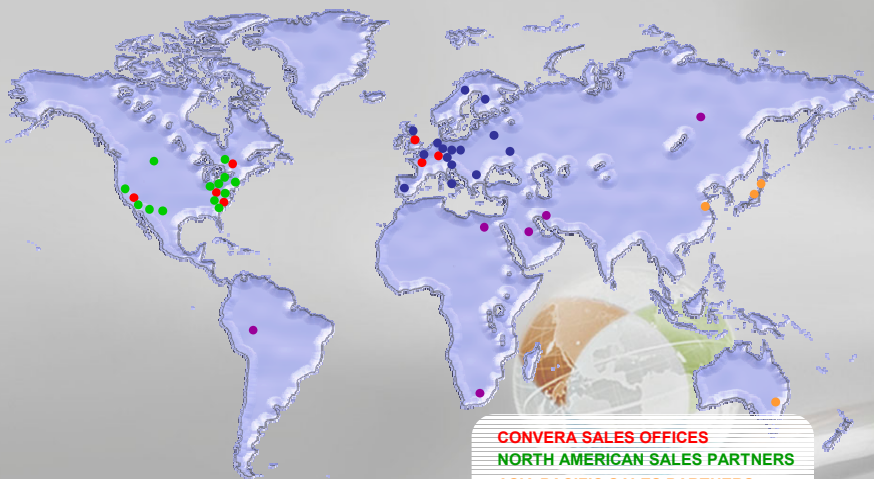


Convera is a leading provider of **enterprise wide index, search and categorization** software products and solutions

- 20+ years of innovation in intelligent information infrastructure
- 250 employees
- 800 customers in 29 countries
- 70+ business partners
- Publicly traded (NASDAQ: CNVR)

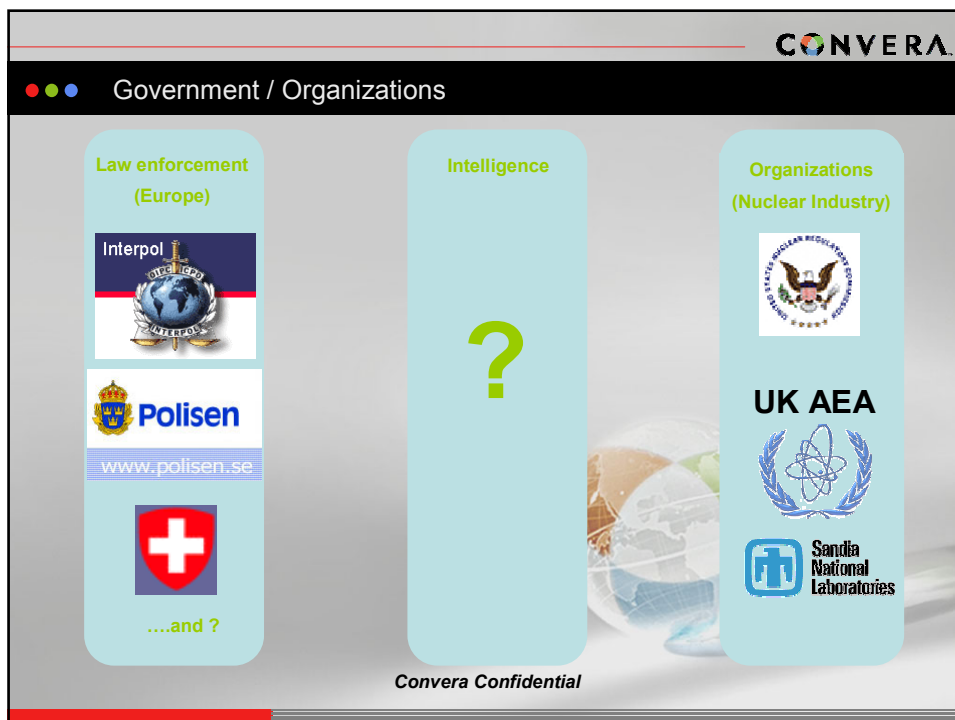
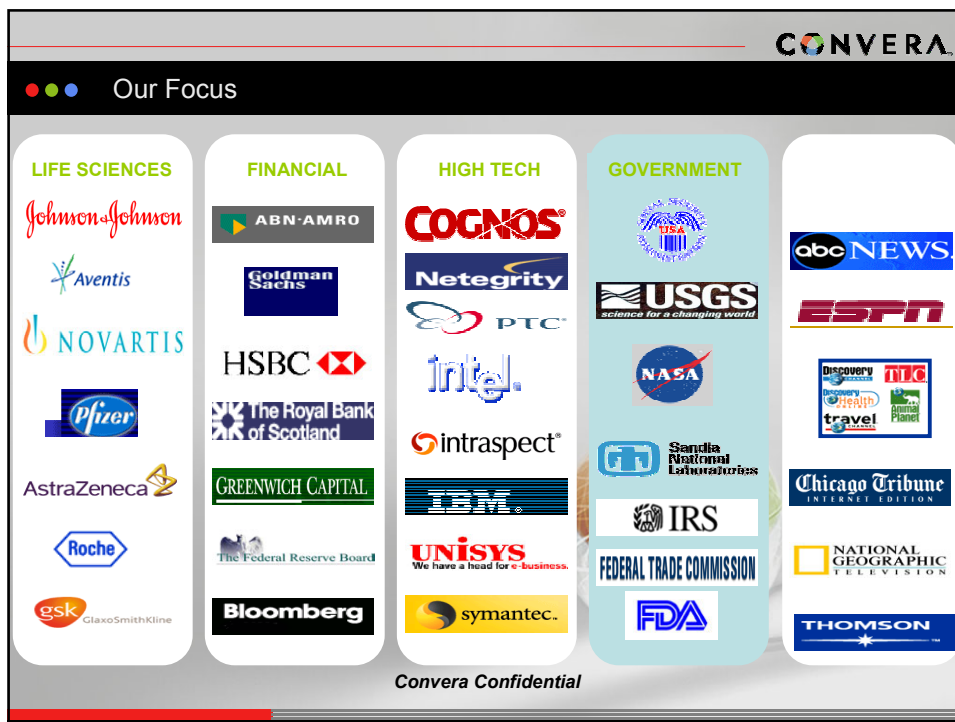
Convera Confidential

Convera & Partner Offices



CONVERA SALES OFFICES
NORTH AMERICAN SALES PARTNERS
ASIA-PACIFIC SALES PARTNERS
EUROPEAN SALES PARTNERS
ROW SALES PARTNERS

Convera Confidential



NRS – CONVERA RetrievalWare Public Accessible

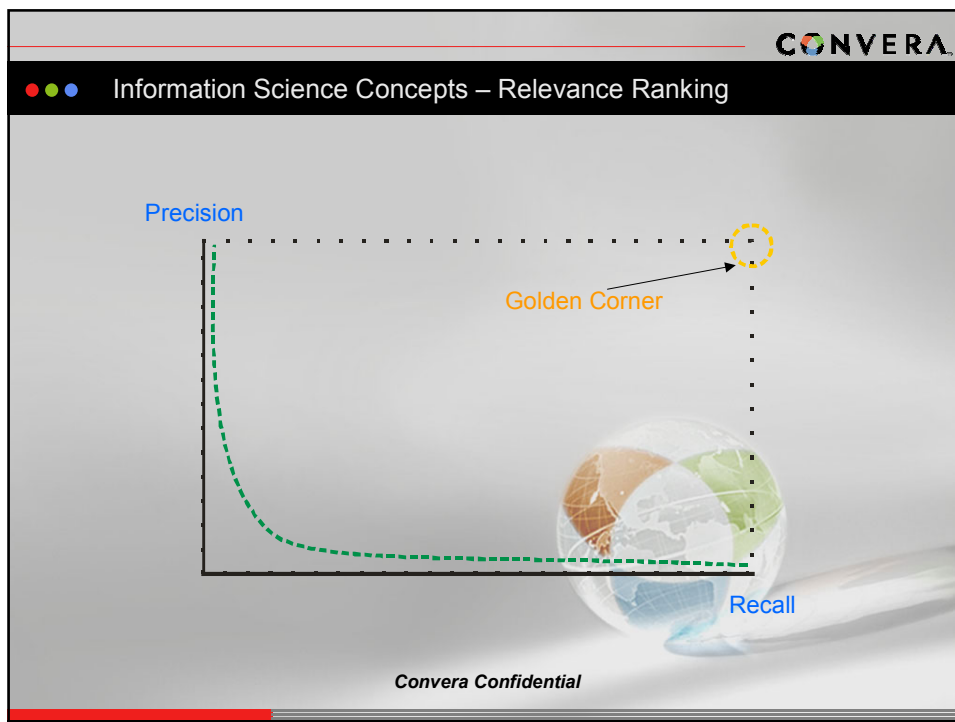
Convera Confidential

Agenda

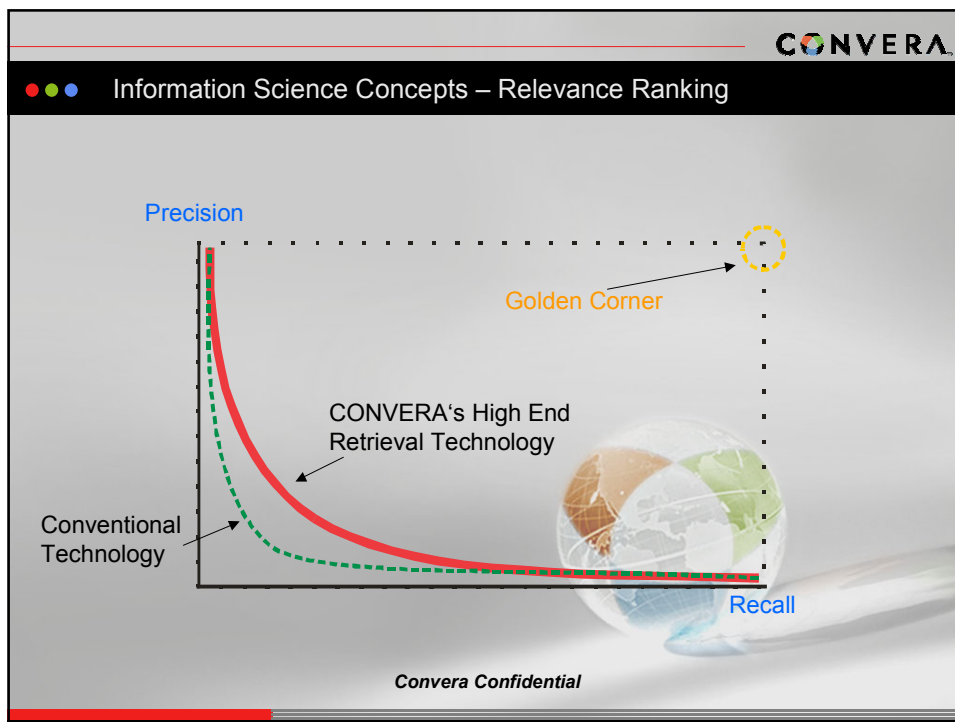
- About CONVERA
- [Introduction](#)
- CONVERA Platform Infrastructure
- CONVERA's Retrieval Technologies
- Visualisation
- Automatic Categorization Technologies
- Dynamic Classification Technologies
- What we do / What's Third Party
- Product demonstration



Convera Confidential



- CONVERA
- Relevance Ranking: Why CONVERA beats!
1. Completeness
 2. Proximity
 3. Hit Density
 4. Semantic Distance
 5. Contextual Evidence
-
- A globe is shown in the bottom right corner of the slide, partially obscured by the list items.
- Convera Confidential



CONVERA

Information Retrieval

1. We know what we know

Example: Today's Weather

A globe graphic is positioned in the bottom right corner of the slide. The text 'Convera Confidential' is at the bottom.

Convera Confidential

Information Retrieval

1. We know what we know

Example: [Today's Weather](#)

2. We know what we don't know

Example: [Winner of 1988 Oscars](#)



Convera Confidential

Information Retrieval

1. We know what we know

Example: [Today's Weather](#)

2. We know what we don't know

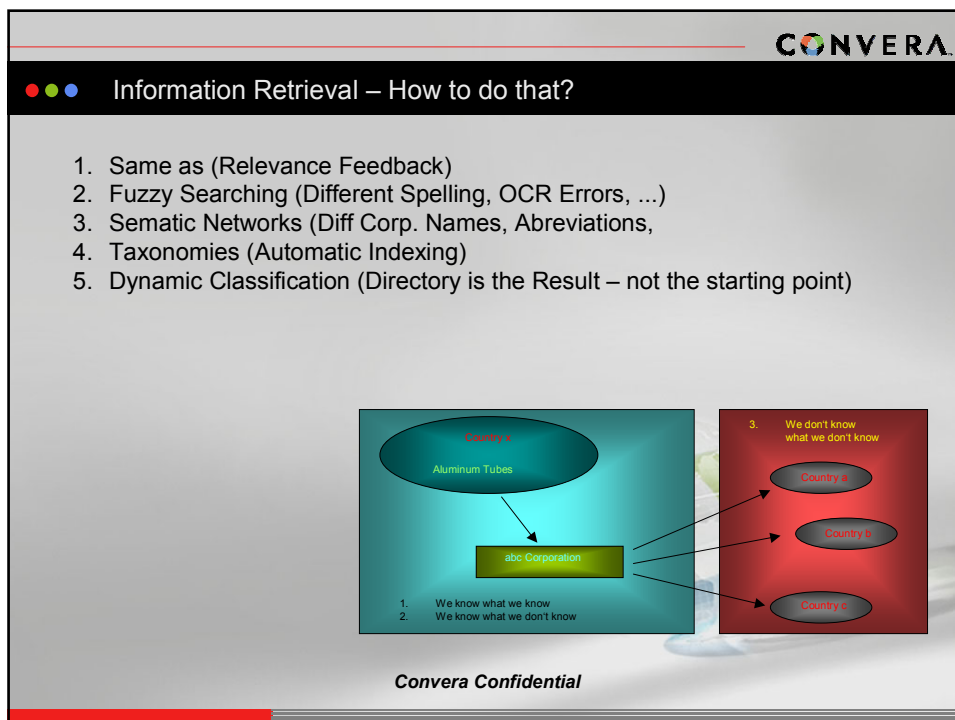
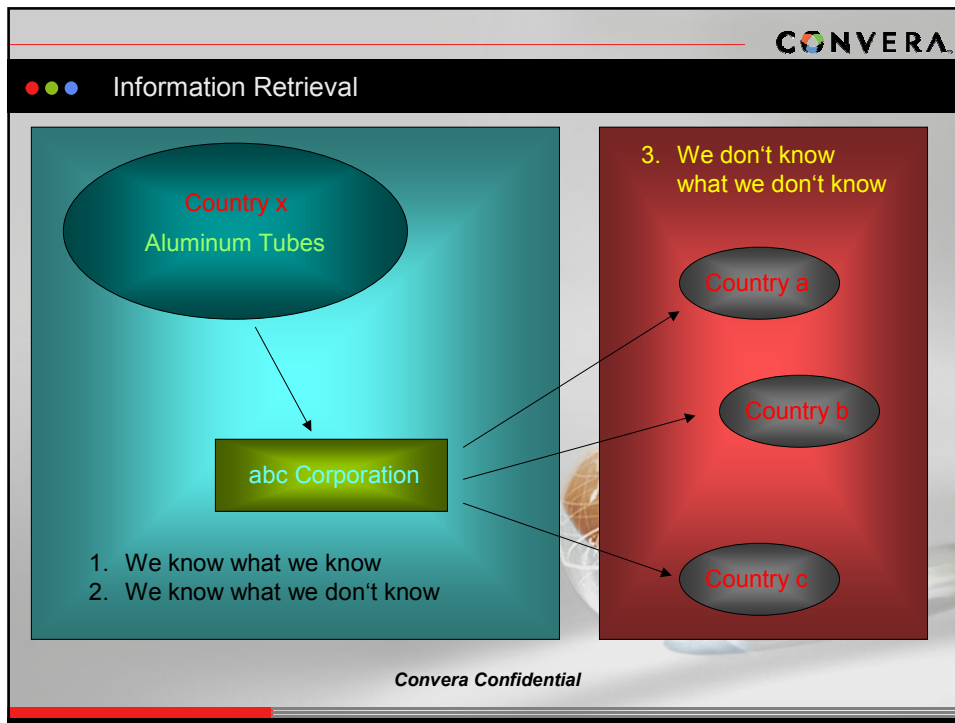
Example: [Winner of 1988 Oscars](#)

3. We don't know what we don't know

Example: [Discovering unknow user of dual purpose Technology](#)



Convera Confidential



Agenda

- About CONVERA
- Introduction
- **CONVERA Platform Infrastructure**
- CONVERA's Retrieval Technologies
- Visualisation
- Automatic Categorization Technologies
- Dynamic Classification Technologies
- What we do / What's Third Party
- Product demonstration



Convera Confidential

Platform Infrastructure



Convera Confidential

Secure, standards-based integration for convenient, single-point of access. B6

Agenda

- About CONVERA
- Introduction
- CONVERA Platform Infrastructure
- [CONVERA's Retrieval Technologies](#)
- Visualisation
- Automatic Categorization Technologies
- Dynamic Classification Technologies
- What we do / What's Third Party
- Product demonstration

*Convera Confidential*

Technologies

- [APRP – Pattern recognition](#)
- List Search Server
- Boolean
- [Concept](#)
- [Language Modules](#)
- [Multi Language Modules](#)
- [Cross Language Modules](#)
- Information Profiling
- [Automatic Categorization](#)
- [Dynamic Personal Classification](#)

*Convera Confidential*

APRP solves that Problem (Mu'ammam al-Qadhafi) better

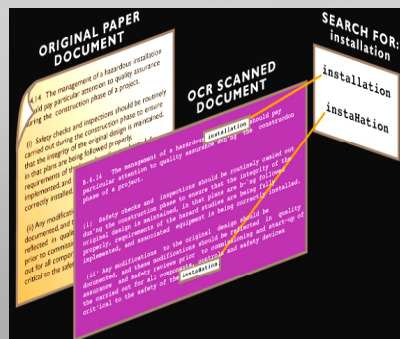
- | | |
|---------------------------|-------------------------|
| 1) Muammam Qaddafi | 17) Moamar Gaddafi |
| 2) Mo'ammam Gadhafi | 18) Mu'ammam Qadhdhafi |
| 3) Muammam Kaddafi | 19) Muammam Khaddafi |
| 4) Muammam Qadhafi | 20) Muammam al-Khaddafi |
| 5) Moammam El Kadhafi | 21) Mu'amar al-Kadafi |
| 6) Muammam Gadafi | 22) Muammam Ghaddafy |
| 7) Mu'ammam al-Qadafi | 23) Muammam Ghadafi |
| 8) Moamer El Kazzafi | 24) Muammam Ghaddafi |
| 9) Moamar al-Gaddafi | 25) Muamar Kaddafi |
| 10) Mu'ammam Al Qathafi | 26) Muammam Quathafi |
| 11) Muammam Al Qathafi | 27) Mohammer Q'udafi |
| 12) Mo'ammam el-Gadhafi | 28) Muammam Gheddafi |
| 13) Moamar El Kadhafi | 29) Muamar Al-Kaddafi |
| 14) Muammam al-Qadhafi | 30) Moammam Khadafy |
| 15) Mu'ammam al-Qadhdhafi | 31) Moammam Qudhafi |
| 16) Mu'ammam Qadafi | 32) Mu'ammam al-Qaddafi |

Convera Confidential

APRP: Why more Accuracy then standard Fuzzy?

BOOT 01000010 01001111 01001111 01010100

BOAT 01000010 01001111 01000001 01010100

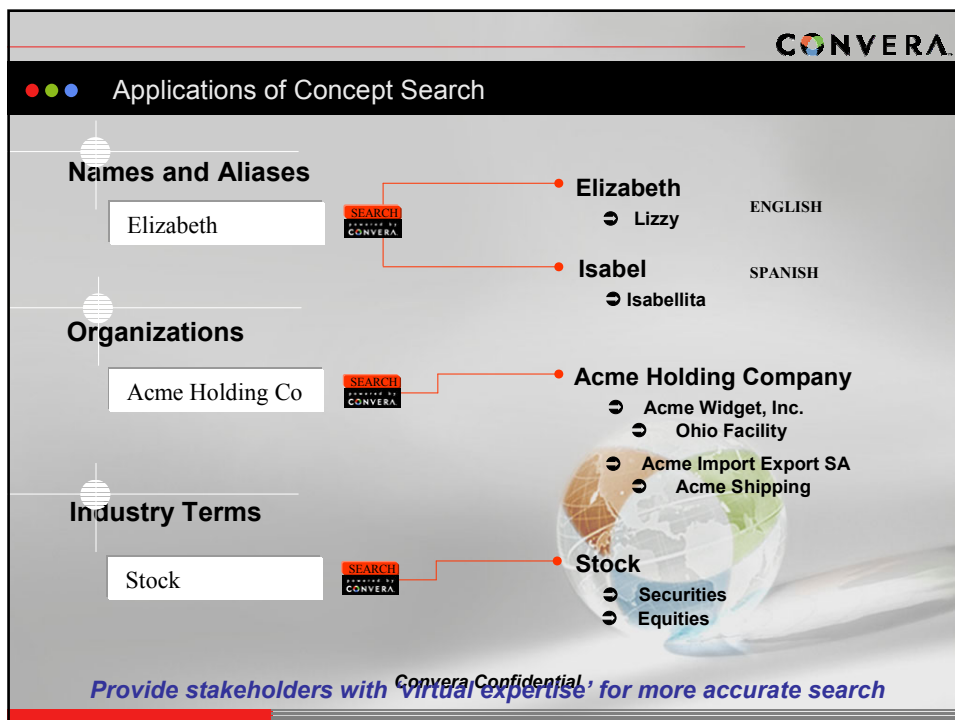
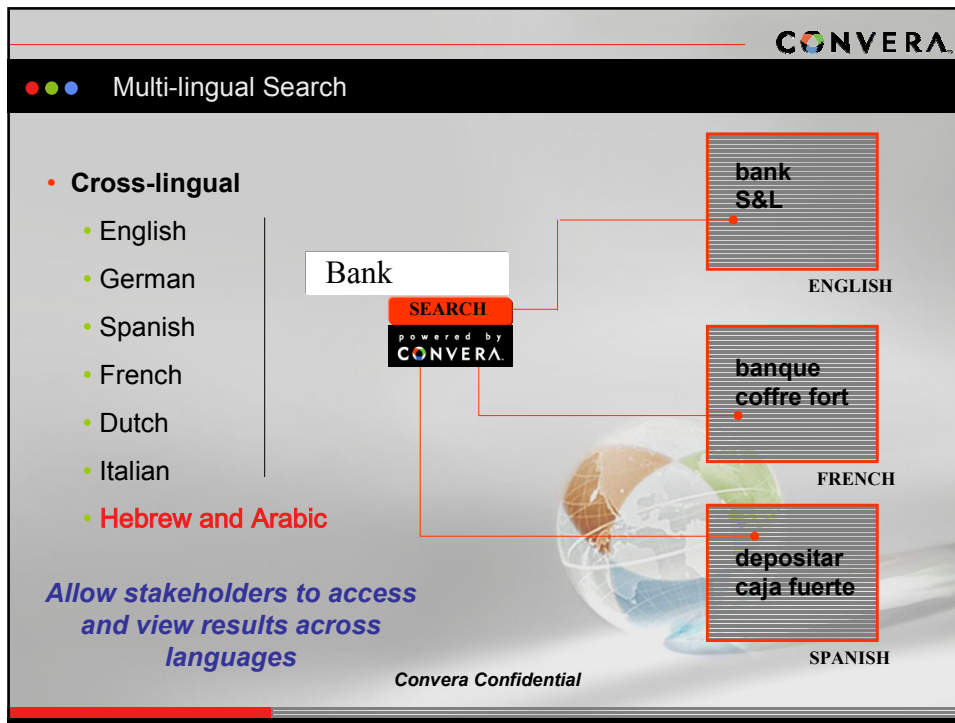


- Overcomes errors, typos, misspellings
- 25% inaccuracy in text is only 10% in binary
- APRP supports multimedia

Sample: Names
Rayn al-Abidin Muamar Hussayn
Rayn al-Abidin Muhammad Husayn

Convera Confidential

22



CONVERA

● ● ● Search Functionality – In the BODY TEXT and in the FIELDS!

- Query by Example (relevance feedback)
- Idiom (Syntactic) Processing
- Adjustable Stop Words
- Exact Phrases
- Date Ranges
- Fielded Searching
- Search Term Weighting
- Logging functions (user)
- Web crawler

- Numeric Range searching
- Multiple Dictionaries / Thesauri
- Recurrent searching (searching hitlist)
- Multiple options for document display
- Automatic categorization
- Dynamic classification
- User profiling
- Relevance Ranking
- Language / Industry Plug-ins
- Cluster displays
- Visualisation aids

Scalability Accuracy
Flexibility

Convera Confidential

CONVERA

● ● ● Third Party Knowledge Discovery Components

Third
Party

CONVERA

Visualization
Tracking
Pattern Detection

Search
Classification
Mapping

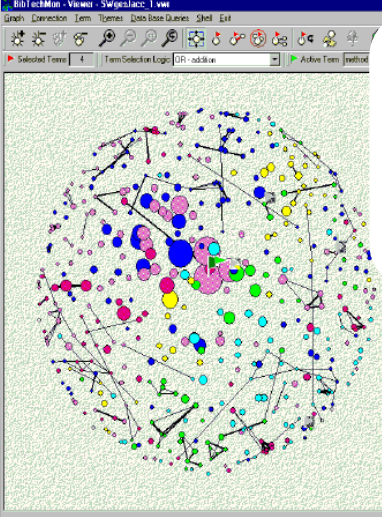
Indexing
Categorization
Entity Extraction

Crawling, Filtering

Convera Confidential

CONVERA

●●● Postprocessing - Clustering

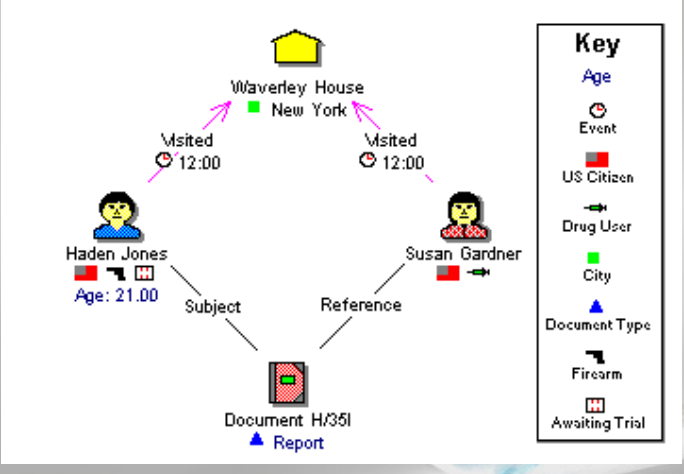


- Clustering
- Statistical Analysis
- Absolute Linking
- Theme Grouping
- etc ...

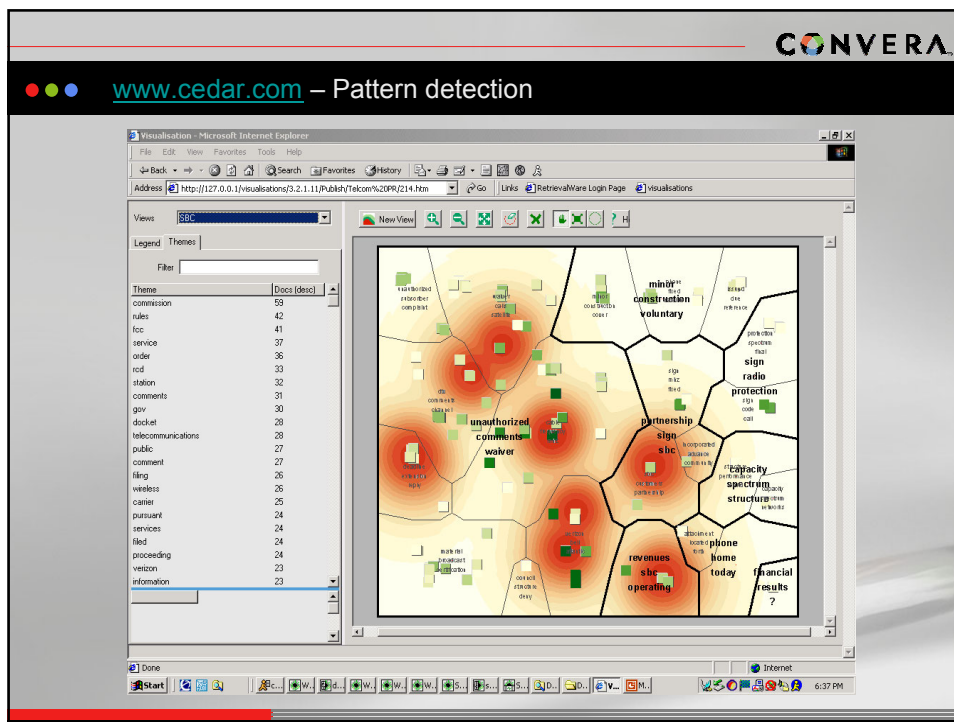
Convera Confidential

CONVERA

●●● www.i2.co.uk - Tracking



Convera Confidential



CONVERA

Agenda

- About CONVERA
- Introduction
- CONVERA Platform Infrastructure
- CONVERA's Retrieval Technologies
- Visualisation
- Automatic Categorization Technologies
- Dynamic Classification Technologies
- What we do / What's Third Party
- Product demonstration

Convera Confidential

• • • The Problem

- Manual indexing is costly and slow
- Traditional Classification is precoordinated
- Hit lists are OK, but somehow inefficient
- Most information is unstructured
- Information structure is irrelevant



Convera Confidential

• • • Ontology

- An **ontology** is a foundation of categories representing a view of the world. An ontology reflects the commonly used and trusted breakdown of categories. For example, the breakdown of news items into categories of 'World', 'Sports', 'Politics', etc. is ontological.



Convera Confidential

Taxonomy

- A **taxonomy** is a hierarchical system describing genera and species. Species derive from a common genus and are hierarchically represented according to their essential characteristics and differences. For example, animals are categorized with the "Taxonomy of Life" which separates mammals from birds and spiders from insects, based on proper features and relative differences. This genus to species nomenclature is highlighted by terminology which moves from generic terms to binomial terms through lexical derivation and compounding.
- A taxonomy doesn't deal with things, but with the essence of things: a taxonomy is based on an ontology.



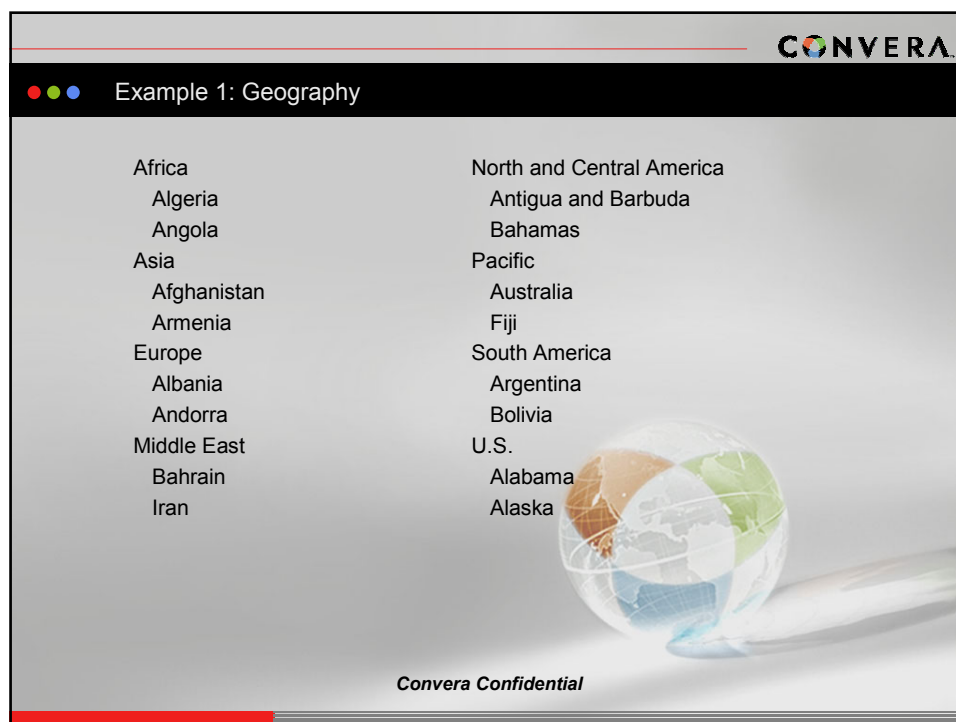
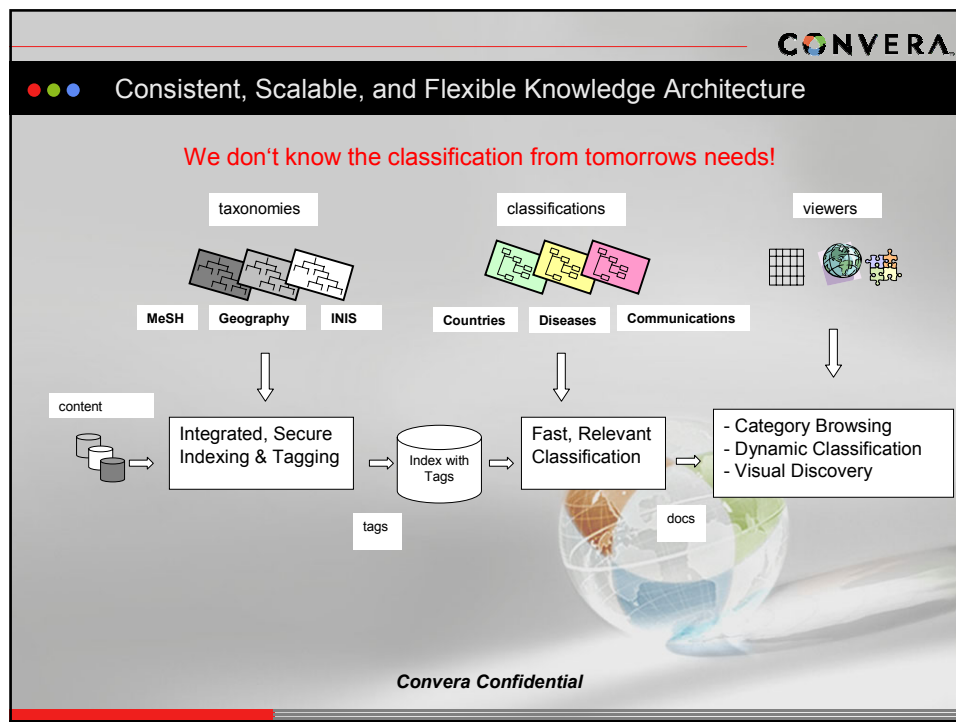
Convera Confidential

Categorization vs. Classification

- | | |
|---|--|
| <ul style="list-style-type: none">• Categorization<ul style="list-style-type: none">• Logical• Taxonomy based• Consistent (Based on cultural fundamentals)• Stable | <ul style="list-style-type: none">• Classification<ul style="list-style-type: none">• Pragmatic• Precordinated• Common sense• Chaotic (Based on best practices)• <u>Individual</u> |
|---|--|



Convera Confidential



Example 2 : Defense

- Defense Communications
 - Satellite Communications
 - Tactical Communications
- Defense Systems
 - Air Defense
 - Antiaircraft Defense Systems
 - Gun Air Defense Systems
 - Antimissile Defense Systems
 - Forward Area Air Defense Systems
 - Terminal Defense
 - Aircraft Defense Systems
 - Antisubmarine Defense Systems
 - Antiswimmer Defense Systems
 - Countermeasures
 - Acoustic Countermeasures



Convera Confidential

Taxonomy Activation

Nairobi

Dar es Salaam

Geography

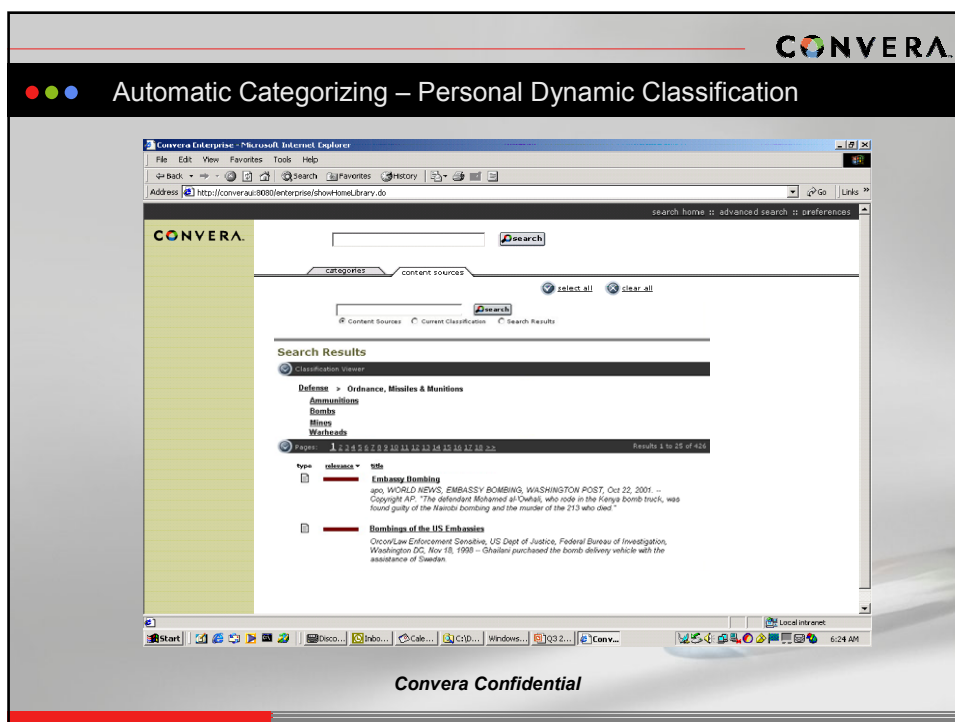
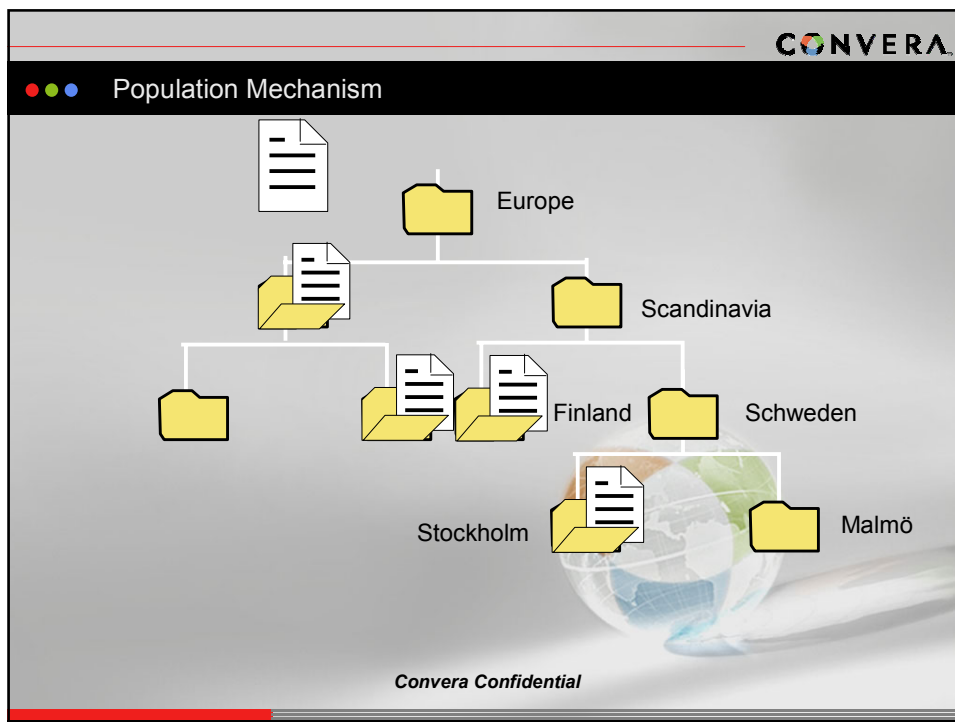
Africa

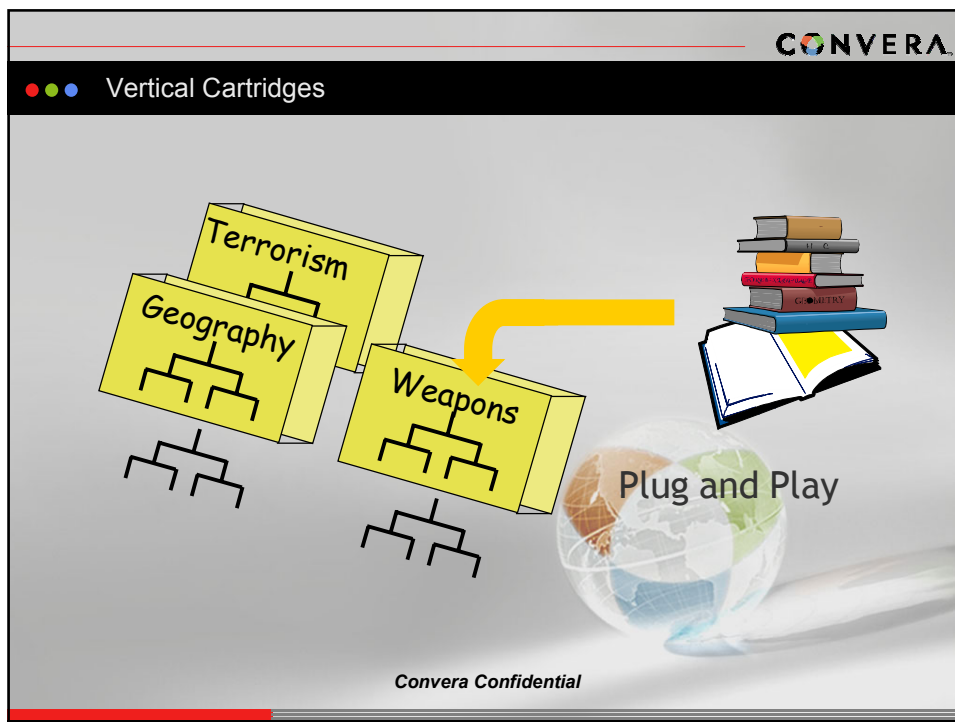
- Algeria
- Angola
- Kenya
- Nairobi
- Tanzania
- Dar es Salaam

Asia

- Afghanistan
- Armenia

Confidential





CONVERA

Visual Discovery

Search UI - Microsoft Internet Explorer

Address: http://localhost:8080/convera/search.jsp

Viewer: Table

Table Viewer Selected

Classification: DTIC_Small

DTIC Small Classification Selected

	Defense Systems (6)	Intelligence and Counterintelligence (79)	Military Operations (13)	War (251)	Ordnance, Missile
Africa (31)	1	5	1	23	1
Asia (77)	1	17	2	52	5
Europe (59)	1	11	3	37	4
Geography (3)		1		2	
Middle East (60)	1	12	2	41	4
South and Central America (17)		5	2	9	1
Pacific (17)				12	3
South America (18)					1
U.S. (34)	2				5

Geo Classification Selected

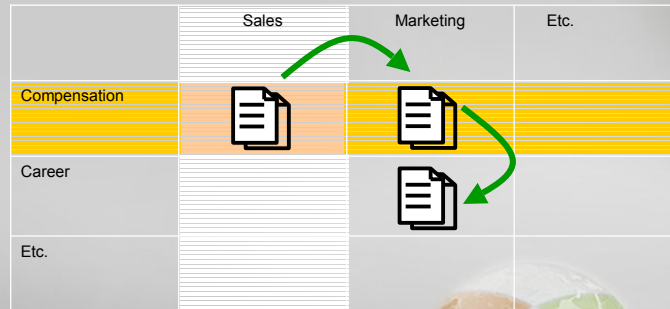
Documents that are related to Military Warfare and Europe

Display: All documents

Local intranet

Convera Confidential

Expand Your Search



- Search is a journey through a multi-dimensional grid of topics
- The ability to visualize all possible combinations at once will save time and increase focus

Convera Confidential

Already Existing Taxonomy Cartridges (Samples)

- Biology
- Chemistry
- Computers
- Electronics
- Finance
- Food Science
- Geography
- Geology
- Health Sciences
- Information Science
- Law
- Mathematics
- MeSH (Medical Subject Headings)
- Military
- Petroleum Natural Gas & Petrochemicals
- Pharmacology
- Physics
- Plastics
- Rubber
- Telecommunications

Convera Confidential

Cartridge Editor enables editing Taxonomy and Dictionary

Taxonomy

Taxonomy Nodes

Synsets

Terms

Dictionary

Cartridge Editor Version 2.02 - File: untitled

File Edit Actions Options ?

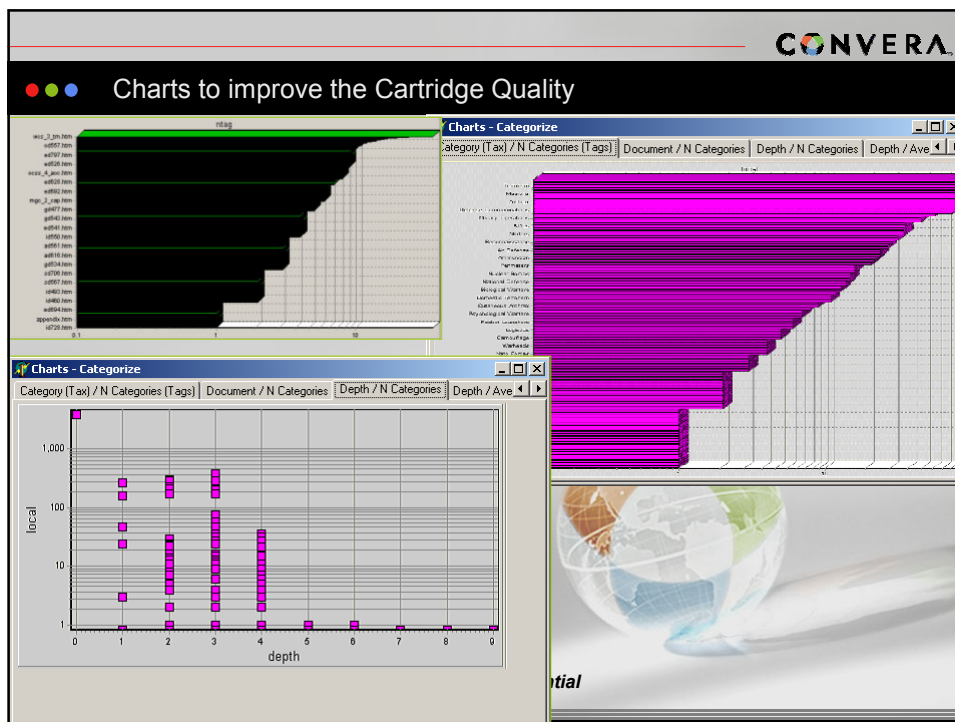
Selected Cartridge: com_convera_mesh Version: 1.0

Convera Confidential

Tables to improve the Cartridge Quality

Tables - Categorize									
Category (Tax) / N Categories (Tags)					Document / N Categories		Category / N Latches		Category / N Ambiguous
	total	LOCAL	depth	ndesc	category	path			
2	382	382	3	1	Missiles	Ordnance, Missiles, and Munitions::Weapons::Missiles			
3	570	293	2	18	Agents	Military Warfare::Agents			
4	333	287	3	3	Terrorism	Military Warfare::Unconventional Warfare::Terrorism			
5	280	280	2	1	National Security	Intelligence and Counterintelligence::National S			
6	630	269	1	16	Intelligence and Counterintelligence	Intelligence and Counterint			
7	231	228	2	5	Missions	Military Operations::Missions			
8	229	211	2	30	Bombs	Ordnance, Missiles, and Munitions::Bombs			
9	209	209	2	4	Battles	Military Warfare::Battles			
10	260	198	3	5	Anthrax	Military Warfare::Agents::Anthrax			
11	815	174	2	58	Weapons	Ordnance, Missiles, and Munitions::Weapons			
12	169	169	3	1	Guns	Ordnance, Missiles, and Munitions::Chemical Ordnance::C			
13	162	160	1	3	Defense Communications	Defense Communications			
14	96	75	3	7	Nuclear Weapons	Ordnance, Missiles, and Munitions::Weapons::t			
15	58	58	3	1	Peacekeeping	Military Operations::Operations Other Than War			
16	396	46	1	34	Military Operations	Military Operations			
17	46	46	3	1	Biological Weapons	Ordnance, Missiles, and Munitions::Weapons::t			
18	36	36	3	1	Persian Gulf War	Military Warfare::Chemical Warfare::Persian Gu			
19	36	36	4	2	Rifles	Ordnance, Missiles, and Munitions::Chemical Ordnance::C			
20	35	35	3	2	Grenades	Ordnance, Missiles, and Munitions::Ammunition::Grenade			
21	35	35	4	1	Assassination	Military Warfare::Unconventional Warfare::Terror			

Convera Confidential



CONVERA

Conclusion

Relate available information to
YOUR
decision-making processes

- Categorize with consistency
- Classify in context

Convera Confidential

The slide features a header with the CONVERA logo and the word "Conclusion". The main content area has a light gray background. In the center, the text "Relate available information to YOUR decision-making processes" is displayed, with "YOUR" in orange and the rest in green. Below this, a white box contains two bullet points: "Categorize with consistency" and "Classify in context". At the bottom center, the text "Convera Confidential" is written. On the right side, there is a graphic of a globe with a magnifying glass over it, symbolizing analysis and decision-making.



Key Differentiators

- The directory is a result, not a starting point.
- Ontologies are real ontologies: conceptual and explicit.



Convera Confidential



Typical Structures

- Geography / Topic
 - Terrorism in Philippines
 - Criminal Law in Texas
 - Domestic Sales
 - Security in Building C
- Horizontal / Vertical
 - Petroleum Business
 - AML Regulations
- Vertical / Vertical
 - Chemical Compounds for Alzheimer



Convera Confidential

Over Defined Context

- Very large computational space
 - “Chemical Compounds in Alzheimer Genomics”
 - -> 8500 diseases
 - -> 1,000 genes
 - -> 30,000 compounds
 - = 255 billion folders
- Reversely proportional number of successfully populated folders
- Can't be done by automatic CLASSIFICATION!



Convera Confidential

What We Do / What's Third Party

What We Do (Some Samples)

- List Search (Batch Mode)
- Information Profiling
- Automatic Categorizing
- Dynamic Classification
- Multi- and Cross Language
- Content Management
- Voice To Text and Automatic Meta Data Generation

Third Party (Some Samples)

- Pre Processing (nCase, etc.)
- Post Processing (Statistics, Facerec, ...)



Convera Confidential



RetrievalWare 8
High End Categorization/Classification

Heinz Bachmann



Convera Confidential Information – The contents of this Convera product presentation are confidential and governed by the NDA between your company and Convera. Such contents are subject to change by Convera at its sole discretion and Convera assumes no obligation to update such contents. Any binding representations, warranties and covenants by Convera shall be exclusively set forth in writing in a contract mutually agreed to and signed by your company and Convera, and such contract shall exclude all other written and oral communications including this presentation.

