

inspiring business decisions

**Business Information the Way you Want it...**

  
**factiva.**  
Dow Jones & Reuters

# Factiva Services

## Content Integration Solutions

- ✓ Factiva Publisher
- ✓ Factiva Search Module
- ✓ Factiva Track Module
- ✓ Factiva Select

**Factiva**  
Factiva.com

## Web-based Services

Dow Jones Interactive  
Reuters Business Briefing

inspiring business decisions

**factiva**<sup>™</sup>  
Dow Jones & Reuters

# Structured, not unstructured, or... lost in the supermarket?



Where do I  
find the  
cornflakes?

# What you need is a pre-defined taxonomy!



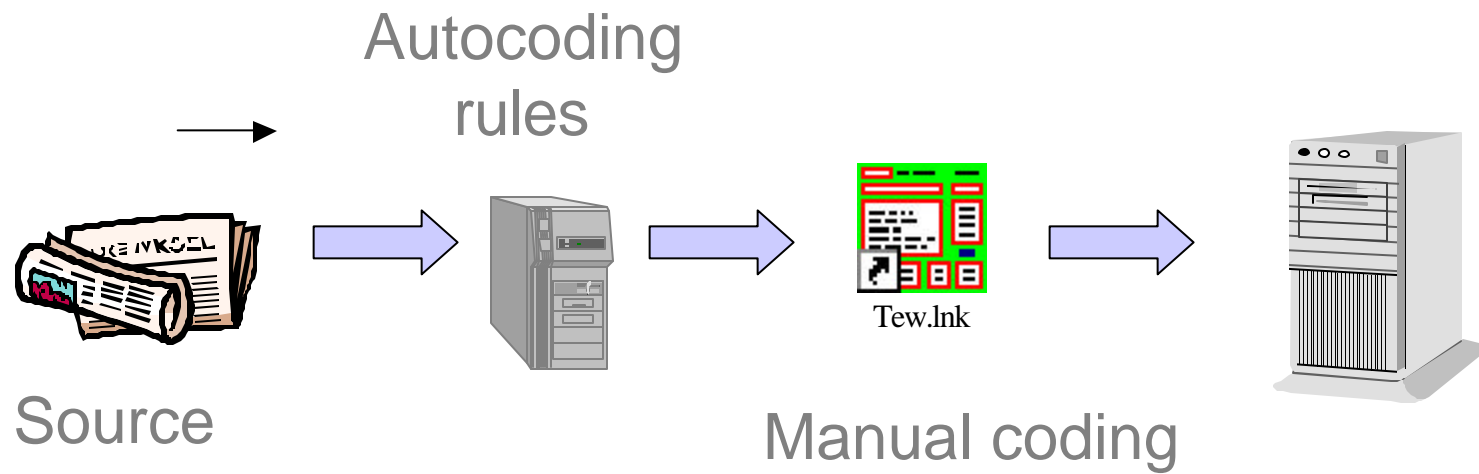
**FOOD PRODUCTS**  
**Breakfast Cereals**  
Oats  
Muesli  
Cornflakes

# Factiva User Case Study: News

## Taxonomy application challenges:

- Many categories: over 1,400 subjects, 300,000 company codes
- High-volume: up to 200,000 stories per day
- Heterogeneous: over 8,000 sources
- Dynamic: changing set of codes, according to what's topical
- Multilingual: 22 content languages

From manual coder....



....to content analyst...

Autocoding rules

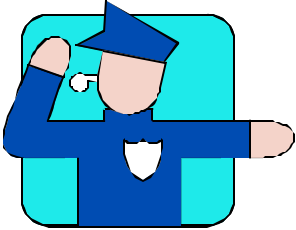


Tew.Ink

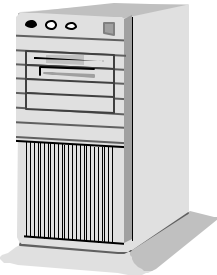
Manual coding



Source



Autocoding analyst

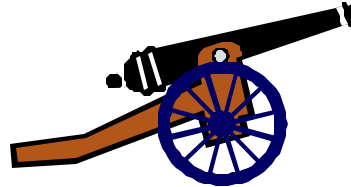


Plus...

another weapon!

Inxight

Autocoding  
rules

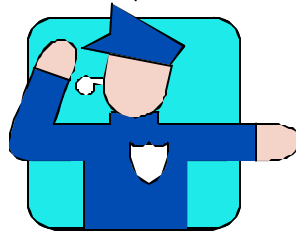
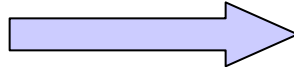


Tew.Ink

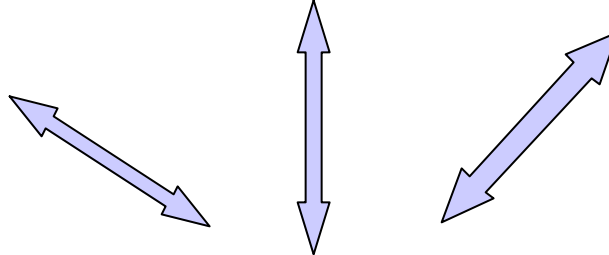
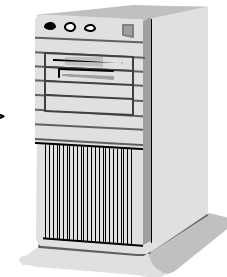
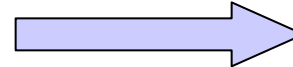
Manual coding



Source



Autocoding  
analyst



inspiring business decisions

**factiva**<sup>™</sup>  
Dow Jones & Reuters

# Factiva User Case Study: News

## Approach:

- Set vendor selection criteria
- Assume example-based “lazy learning” for subject categories, so avoid rule-writing/training/retraining
- Use named-entity extraction to enhance company coding
- Expect you will still need humans (target 45 percent autoloading)

# Factiva User Case Study: News

## Testing, testing, testing:

- Entity-finding was instantly very effective
- For subject coding, the initial training set was not adequate  
Create a fresh, manually coded training set – a lot of work (and it still wasn't perfect!)
- Real-world use shows an uneven distribution of codes, so focus on areas which give maximum Return on Investment

# Factiva User Case Study: News

## Results:

- Over 70 percent of stories tested were autoloading via Inxight alone; more using additional techniques
- Careful, controlled attention to the training set is very important
- High-end editors need excellent diagnostic tools
- The training data is a “starter set” for other implementations

## What next?

Doing it to ourselves: tackling Factiva's internal knowledge base:

- Apply Factiva subject codes using Inxight Categorizer
- Apply Factiva company codes using Inxight Thing Finder
- Apply Factiva internal taxonomy using Inxight Categorizer

# Factiva Corporate Intranet – personal-page view

**open portal DEMO** \*Sitemap \*News \*People | Others: GO TO: inight

**Summary** ThingView

USA: N.American pork producers seek open Taiwan markets.

CHICAGO, May 2 (Reuter) - North American pork producers on Friday called on Taiwan to open its markets for pork imports, with the island's own hog herd devastated by an outbreak of foot and mouth disease.

The U.S.-based National Pork Producers Council (NPPC) hosted a meeting in Chicago with officials of the Canadian Pork Council, the Mexican Pork Council and the Mexico-based National Commission for Pork Products (CONAPOR).

**ThingSearch pointing to: Reuters**

COMPANY/ORGANIZATION	PEOPLE	PLACE
ABM Industries	Abdul Pillane	Adelaide
ABM Industries Inc	Adelaide	Africa
ABN AMRO Holding N.V.	Adrian	Alamo
ATT	Adrian Edwards Hanoi	Alaska
Action Party	Ahmed Abdel-Rahman	Algeria



# An Evaluation of Modern Categorization Systems...

Questions?