

Living Documents - Semantic Tagging of User-Generated Content



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Aljoscha Burchardt, Prof. Iryna Gurevych, UKP Lab

What is this text about?



CHICAGO, Oct 29 - Kraft Foods Inc and Kellogg Co posted better-than-expected third-quarter profits on Wednesday as price increases and new products helped lift sales in a weak economy. Kraft also stood by its forecasts for 2008 earnings before one-time items as well as for 2009 net income, while Kellogg said its profit this year should hit the high end of its previous targeted range. Both Kraft, the largest North American food maker, and Kellogg, the world's largest cereal company, have taken steps to cut costs and put more money into advertising. Both have also bolstered new product development to attract consumers even as rising commodity costs pushed them to raise prices. Commodities like wheat and energy have become less expensive in recent months, but food companies may not see a big benefit until next year, in part because they lock in their costs months ahead. Kraft, which makes Oreo cookies, Tang breakfast drink and Oscar Mayer hot dogs, reported a profit of 45 cents a share before one-time items, a penny above what analysts polled by Reuters Estimates had expected. The company hiked prices on products, leading to a 0.9 percent drop in volume. However, that key result was still better than the company had expected. Analysts are watching to see how much consumers cut back on buying branded products in the face of rising food prices and a slumping economy. Kraft sales rose 19.4 percent to \$10.46 billion. Organic sales, which exclude the impact of currency, acquisitions and divestitures, rose 7.1 percent due to higher pricing. Over the past several years, Kraft has closed factories, cut jobs and divested brands to focus on areas like cookies and crackers, pizza and healthier foods. Shares of Kraft rose 1.8 percent to \$29.39 in premarket trading from Tuesday's closing price of \$28.88 on the New York Stock Exchange, while Kellogg stock was not active. Kellogg, the maker of Rice Krispies and Eggo waffles, said net income rose to 89 cents a share from 76 cents a year earlier. The results were far better than the 80 cents analysts had forecast.

Highlighting Keyphrases Speeds Up Understanding



CHICAGO, Oct 29 - **Kraft Foods Inc** and **Kellogg Co** posted **better-than-expected third-quarter profits** on Wednesday as **price increases** and new products helped **lift sales** in a **weak economy**. **Kraft** also stood by its forecasts for 2008 earnings before one-time items as well as for 2009 net income, while **Kellogg** said its profit this year should hit the high end of its previous targeted range. Both **Kraft**, the **largest North American food maker**, and **Kellogg**, the **world's largest cereal company**, have taken steps to cut costs and put more money into advertising. Both have also bolstered new product development to attract consumers even as rising commodity costs pushed them to raise prices. Commodities like wheat and energy have become less expensive in recent months, but food companies may not see a big benefit until next year, in part because they lock in their costs months ahead. Kraft, which makes Oreo cookies, Tang breakfast drink and Oscar Mayer hot dogs, reported a profit of 45 cents a share before one-time items, a penny above what analysts polled by Reuters Estimates had expected. The company hiked prices on products, leading to a 0.9 percent drop in volume. However, that key result was still better than the company had expected. Analysts are watching to see how much consumers cut back on buying branded products in the face of rising food prices and a slumping economy. **Kraft sales** rose 19.4 percent to \$10.46 billion. Organic sales, which exclude the impact of currency, acquisitions and divestitures, rose 7.1 percent due to higher pricing. Over the past several years, **Kraft** has closed factories, cut jobs and divested brands to focus on areas like cookies and crackers, pizza and healthier foods. Shares of **Kraft** rose 1.8 percent to \$29.39 in premarket trading from Tuesday's closing price of \$28.88 on the New York Stock Exchange, while **Kellogg** stock was not active. **Kellogg**, the maker of Rice Krispies and Eggo waffles, said net income rose to 89 cents a share from 76 cents a year earlier. The results were far better than the 80 cents analysts had forecast.



- **The High Cost of Not Finding Information** (IDC whitepaper 2003)
 - An IT manager for one Fortune 1000 company: almost 25% of documents on the corporate intranet are not indexed.
- **But don't we have semantic search?**
- **Yes, but one size does not fit all. You may want to:**
 - Browse through the intranet to analyze your companies strengths
 - Cluster and label documents that belong together
 - Ensure corporate standards
 - Highlight important information in longer documents

Information Has to Be Structured



TECHNISCHE
UNIVERSITÄT
DARMSTADT





Why Not Have the Users Add Tags?

- **Time efficiency**
 - **Motivation**
 - **It's not always easy to find the relevant terms in a document**
 - **Printer** might suffice in a law firm
 - **Poster printer** might be appropriate in a PR office
 - **Poster printer model XY** in a hardware store
- ➔ Computers using semantic technology can do better**

What UKP Lab Can Offer



- **UKP offers automatic key-phrases extraction based on state-of-the-art language technology**
- **Current approaches based purely on word frequency („Google principle“) are problematic for intranets**
 - Rare key words tend to be ignored
 - Insufficient precision and recall (missing redundancy in document set)
- **UKP's technology**
 - Identifies key **phrases** rather than only words, e.g., *meeting with X*
 - Relevance of key phrases is assessed on the basis of **semantic relations**
- **Key principle:** represent texts (internally) as semantic graphs

Visualization of a Text as Semantic Graph



MMAX2 1.12 D:\Programme\eclipse\workspace\lexgraphs\output\mmax\article_4.txt.mmax [modified]

File Settings Display Tools Plugins Info Show ML Panel

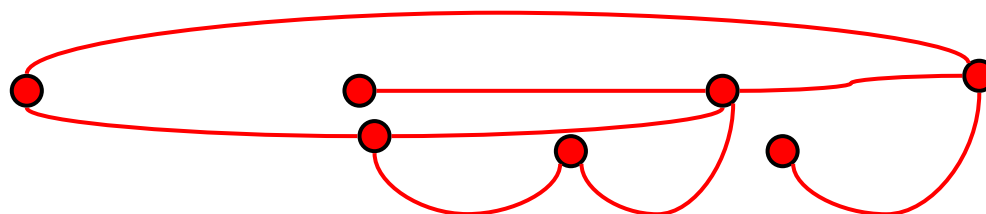
IWF : Weltwirtschaft robust - Deutschland hinten WASHINGTON - Die Weltwirtschaft wächst in diesem Jahr trotz einer leichten Abschwächung weiter robust - allerdings muss sich Deutschland im internationalen Wachstumsvergleich mit einem der hinteren Plätze begnügen . Nach Einschätzung des Internationalen Währungsfonds (IWF) ist 2007 weltweit mit einem Wachstum von 4,9 Prozent zu rechnen , nach 5,4 Prozent 2006 . Während die USA 2007 ein Plus von 2,2 Prozent erwarten dürfe , blieben Deutschland und Italien mit je 1,8 Prozent Schlusslicht unter den großen Industrieländern . Allerdings hob der IWF unter den entwickelten Ländern für dieses Jahr seine Prognose für Deutschland mit 0,6 Prozentpunkten so stark an wie sonst nur noch für Spanien . Deutschland braucht mehr Wettbewerb im privaten Sektor , um die Produktivität zu erhöhen " , riet der federführende Ökonom des Weltwirtschaftsausblicks , Timothy Callen . Obwohl schon zahlreiche Reformen angegangen worden seien , müssten zusätzliche folgen , zusammen mit einer weiteren Begrenzung der Staatsausgaben . Weltweit rechnet die Finanzinstitution für 2008 abermals mit einem Plus vom 4,9 Prozent . Für die USA als größte Volkswirtschaft erwartet der IWF im kommenden Jahr mit 2,8 Prozent wieder ein höheres Wachstumstempo . Deutschland sinkt der Fonds indes mit 1,9 Prozent erneut unter den Schlusslichtern , auf gleicher Höhe mit Japan , aber vor Italien , Großbritannien werde um 2,7 Prozent zulegen , Frankreich um 2,4 Prozent . Der IWF ist mit seinem Ausblick für dieses Jahr pessimistischer für Deutschland als deutsche Konjunkturforscher und Banken-Volkswirte . Diese erwarten meist ein Plus von zwei Prozent und mehr . Während sich Deutschland im vergangenen Jahr mit 2,7 Prozent Plus laut IWF dank starker Exportzuwächse und hoher Investitionen als Wachstumslokomotive im Euro-Raum hervorgetan habe , mache sich nun eine erwartete Konsumabkühlung durch die Erhöhung der Mehrwertsteuer bemerkbar . Dieser Schritt werde auch die Inflation anheizen . Insgesamt bewerten die Experten des Währungsfonds die Wirtschaftsentwicklung im Euro-Raum nach dem 1 . Quartal 2007 als soweit " nachhaltig " . In Europa sieht der IWF Großbritannien mit 2,9 Prozent als Spitzenreiter beim Wirtschaftswachstum . Alles in allem bewertet der Fonds die Aussichten für die Weltwirtschaft positiver als noch vor sechs Monaten . Auswirkungen der Krise am Immobilienmarkt der USA seien soweit begrenzt und Inflationsrisiken hätten sich abgeschwächt . " Die Risiken für die Wachstumsaussichten sind weniger bedrohlich als im September 2006 " , hieß es .

Example Semantic Graph



Michael Schumacher won his last Grand-Prix for Ferrari in 2001. He was the only driver to win 4 championships in a row.

Michael Schumacher won his last Grand-Prix for Ferrari in 2001. He was the only driver to win 4 championships in a row.



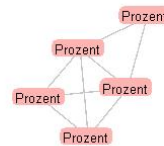
Graph Clustering



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Graph Prefuse Consumer - muh.txt

Finanzinstitution



IWF

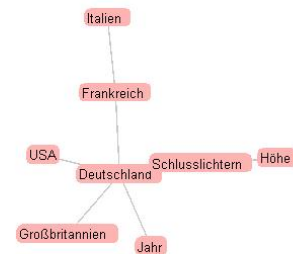
Fonds

Japan

Volkswirtschaft

Wachstumstempo

Plus



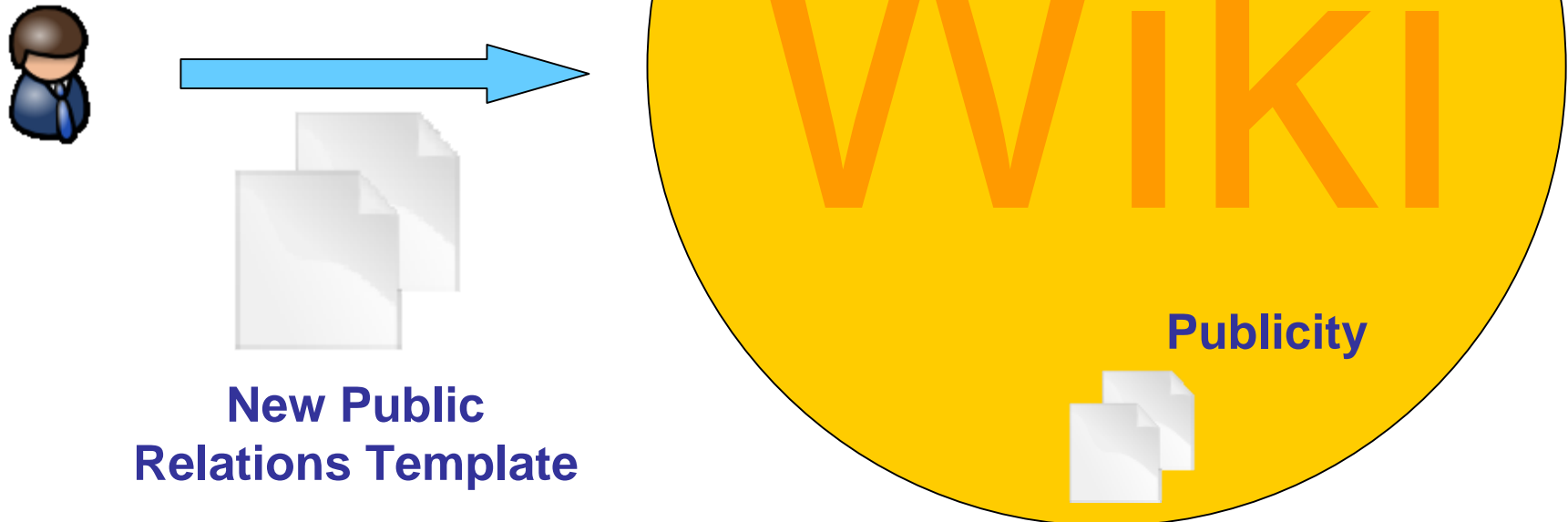
Prototype Application: The self-organizing Wiki

(Work in Progress at UKP)



TECHNISCHE
UNIVERSITÄT
DARMSTADT

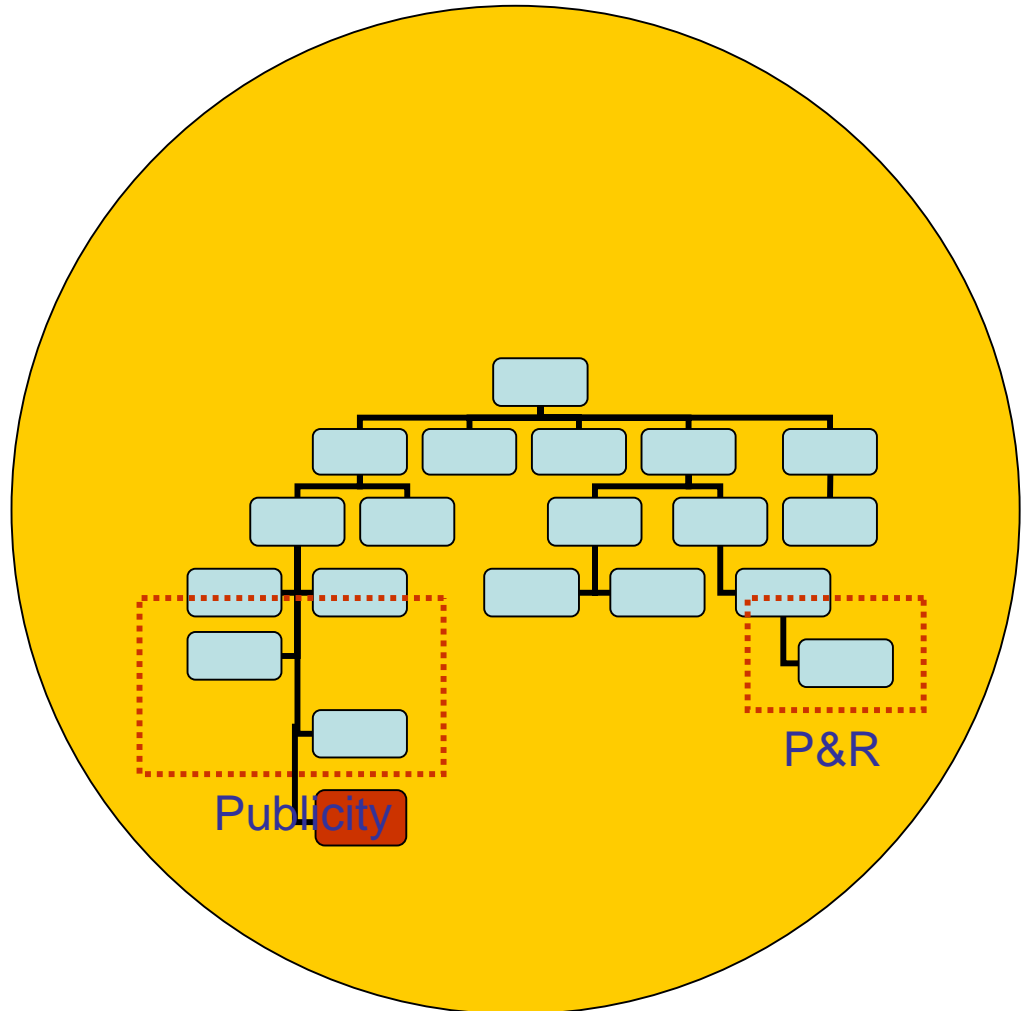
Problem – Filing New Information



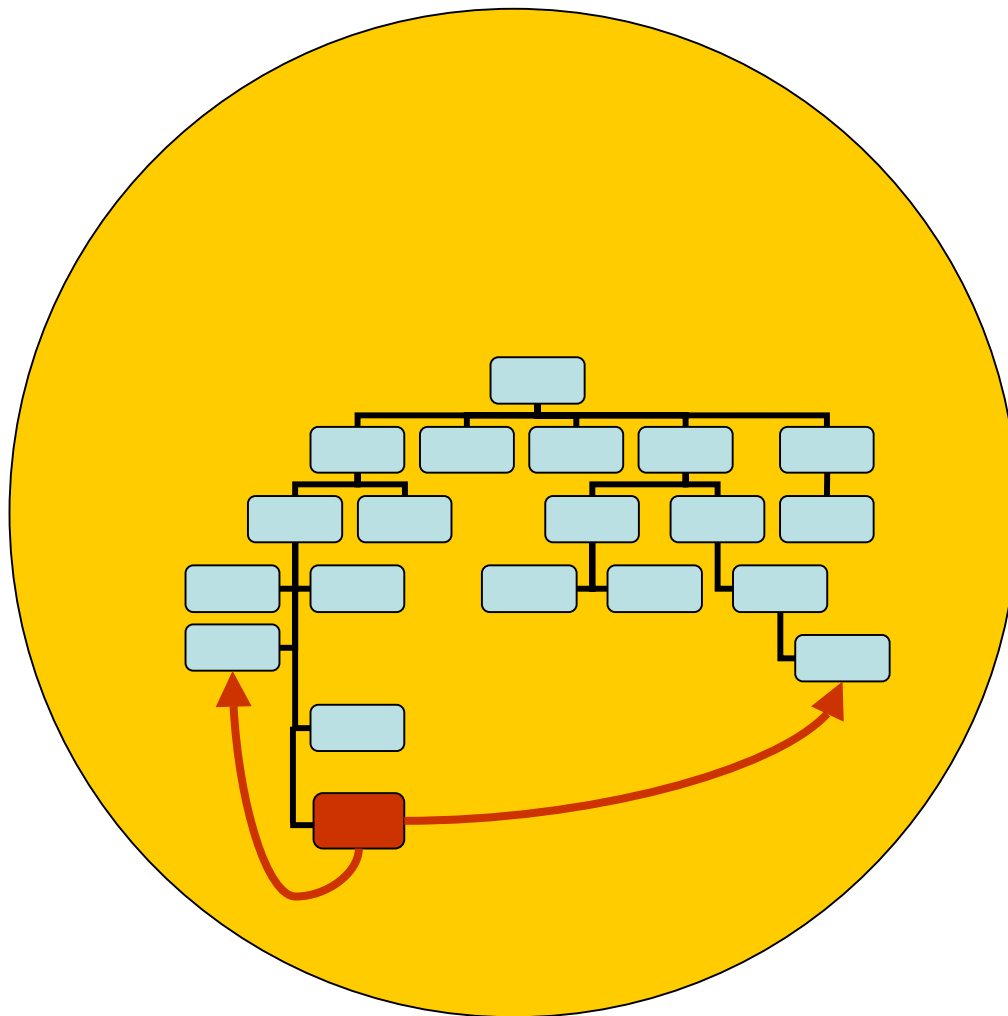
Solution – Automatic Suggestion of Topic



Public Relation
Templates



Automatic Linking



Summary



- **UKP's technology**
 - Automatically identifies relevant **key phrases...**
 - ... using **semantic word/concept relations**
 - Is specialized on user-generated (2.0) content
- **Potential applications in industry**
 - **Information management**
(tagging, categorization)
 - **Optimizing communication** and collaboration
(clustering, highlighting information)
 - **Quality control** of ECM
(browsing, searching)
- **Potential business partners**
 - IT companies
 - Companies with an own IT department



Requirements Concerning the Transfer Business

- We are looking for business partners
- We want to achieve: R&D cooperation, commercialization or licensing



Ubiquitous Knowledge Processing Lab
<http://www.ukp.tu-darmstadt.de>