

Informationsextraktion

Methoden und Anwendungen

Vortrag für den Darmstädter Ontologenkreis

von Melanie Siegel

8. Februar 2017

Worum geht es eigentlich?

Data Mining

- strukturierte Daten
- meist Zahlenformate, Tabellen

Text Mining

- unstrukturierte Daten, Text
- Extraktion von Schlüsselwörtern
- Statistik über Schlüsselwörter in den Texten

Informationsextraktion

- unstrukturierte Daten, Text
- Extraktion der relevanten Information
- Information ist vernetzt
- Darstellung in maschinenlesbarem Format
- Strukturierten Zugriff auf dieses Format ermöglichen

Welche Information?

- Das variiert je nach Aufgabe!
- z.B. Personen, Ereignisse, Fehler, Produkte, Meinungen, ...
- semantische Analyse = vernetztes Wissen
- Vernetzung im Sprachausdruck
 - Kollokationen
 - Skopus
 - Argumentstrukturen
 - kompositionale Semantik
- Vernetzung nach außerhalb
 - Sprache -> Welt
 - Ontologien (Wörter -> Ontologie)
 - QA: Fragen an Daten

Beispiel: Fußball

31 May 2002, 19:31 Seoul – Tokyo FRANCE 0 : 1 SENEGAL by FIFAworldcup.com

FIFA World Cup' debutantes Senegal kicked off the 2002 edition of the world's greatest sporting event by shocking holders France 0 : 1 in Seoul, Korea through a first-half goal from Papa Bouba Diop. The French-speaking African side came out bravely, and behind the dangerous El Hadji Diouf, proved that they were not going to be intimidated. France increased their pressure but their finishing failed to break down a solid Senegalese defence. Considering that France hit the woodwork twice, Senegal coach Bruno Metsu could not be more pleased about the end result of the long-awaited match. Never has the loss of Zidane's magic wand been felt more cruelly by the French. Senegal did not seem intimidated at all by the FIFA World Cup holders in the early going, and El Hadji Diouf created the first chance for Senegal when he beat Marcel Desailly to the by-line and laid the ball back for Khalilou Fadiga, whose low shot rolled right to France goalkeeper Fabien Barthez. Shortly after France had a half-chance, when a Barthez punt was knocked on to a streaking David Trezeguet behind the Senegal defence. But the Juventus striker's first touch let him down. Trezeguet's touch failed him again minutes later as he struck the right post with a powerful, curling shot from the top of the penalty area (24'). Senegal shocked France with the tournament's first goal at the half-hour mark, just when it seemed that the French offside trap had Diouf in fits. The Senegalese striker rounded Frank Leboeuf and laid a dangerous ball across the face of the goal. It took deflections from Emmanuel Petit and Barthez before sitting up wonderfully for Papa Bouba Diop, who easily knocked the ball into the open net from his backside (0:1, 30'). After giving up the goal, France finished out the half by creating a flurry of dangerous situations, but they never managed to truly test Sylva. France opened the second half with a string of chances. They went close when Lilian Thuram served a cross from the right side, and Trezeguet's glancing header couldn't stay inside the left post (55'). France missed an even better opportunity soon after. Sylvain Wiltord served across from the right side, and Henry couldn't keep his header down, sailing it over an open goal. Then it was Youri Djorkaeff's turn to have a go. His shot from 25 metres was stopped by Sylva, who had to scramble to punch the rebound over the by-line for a corner. Vieira then rose up to meet a corner from the right side and sent a powerful header on goal, forcing Sylva to make a nice diving save to his right (58').

Zeit und Datum

31 May 2002, 19:31 Seoul – Tokyo FRANCE 0 : 1 SENEGAL by FIFAworldcup.com

FIFA World Cup' debutantes Senegal kicked off the **2002** edition of the world's greatest sporting event by shocking holders France 0 : 1 in Seoul, Korea through a first-half goal from Papa Bouba Diop. The French-speaking African side came out bravely, and behind the dangerous El Hadji Diouf, proved that they were not going to be intimidated. France increased their pressure but their finishing failed to break down a solid Senegalese defence. Considering that France hit the woodwork twice, Senegal coach Bruno Metsu could not be more pleased about the end result of the long-awaited match. Never has the loss of Zidane's magic wand been felt more cruelly by the French. Senegal did not seem intimidated at all by the FIFA World Cup holders in the early going, and El Hadji Diouf created the first chance for Senegal when he beat Marcel Desailly to the by-line and laid the ball back for Khalilou Fadiga, whose low shot rolled right to France goalkeeper Fabien Barthez. Shortly after France had a half-chance, when a Barthez punt was knocked on to a streaking David Trezeguet behind the Senegal defence. But the Juventus striker's first touch let him down. Trezeguet's touch failed him again minutes later as he struck the right post with a powerful, curling shot from the top of the penalty area (24'). Senegal shocked France with the tournament's first goal at the half-hour mark, just when it seemed that the French offside trap had Diouf in fits. The Senegalese striker rounded Frank Leboeuf and laid a dangerous ball across the face of the goal. It took deflections from Emmanuel Petit and Barthez before sitting up wonderfully for Papa Bouba Diop, who easily knocked the ball into the open net from his backside (0:1, 30'). After giving up the goal, France finished out the half by creating a flurry of dangerous situations, but they never managed to truly test Sylva. France opened the second half with a string of chances. They went close when Lilian Thuram served a cross from the right side, and Trezeguet's glancing header couldn't stay inside the left post (55'). France missed an even better opportunity soon after. Sylvain Wiltord served across from the right side, and Henry couldn't keep his header down, sailing it over an open goal. Then it was Youri Djorkaeff's turn to have a go. His shot from 25 metres was stopped by Sylva, who had to scramble to punch the rebound over the by-line for a corner. Vieira then rose up to meet a corner from the right side and sent a powerful header on goal, forcing Sylva to make a nice diving save to his right (58').

Personen und Orte

31 May 2002, 19:31 Seoul – Tokyo FRANCE 0 : 1 SENEGAL by FIFAworldcup.com

FIFA World Cup' debutantes **Senegal** kicked off the **2002** edition of the world's greatest sporting event by shocking holders **France** 0 : 1 in **Seoul, Korea** through a first-half goal from **Papa Bouba Diop**. The French-speaking African side came out bravely, and behind the dangerous **El Hadji Diouf**, proved that they were not going to be intimidated. **France** increased their pressure but their finishing failed to break down a solid **Senegalese** defence. Considering that **France** hit the woodwork twice, **Senegal** coach **Bruno Metsu** could not be more pleased about the end result of the long-awaited match. Never has the loss of Zidane's magic wand been felt more cruelly by the **French**. **Senegal** did not seem intimidated at all by the FIFA World Cup holders in the early going, and El Hadji Diouf created the first chance for **Senegal** when he beat **Marcel Desailly** to the by-line and laid the ball back for **Khalilou Fadiga**, whose low shot rolled right to France goalkeeper **Fabien Barthez**. Shortly after **France** had a half-chance, when a **Barthez** punt was knocked on to a streaking David Trezeguet behind the Senegal defence. But the Juventus striker's first touch let him down. Trezeguet's touch failed him again minutes later as he struck the right post with a powerful, curling shot from the top of the penalty area (24'). Senegal shocked France with the tournament's first goal at the half-hour mark, just when it seemed that the French offside trap had Diouf in fits. The Senegalese striker rounded Frank Leboeuf and laid a dangerous ball across the face of the goal. It took deflections from Emmanuel Petit and Barthez before sitting up wonderfully for Papa Bouba Diop, who easily knocked the ball into the open net from his backside (0:1, 30'). After giving up the goal, France finished out the half by creating a flurry of dangerous situations, but they never managed to truly test Sylva. France opened the second half with a string of chances. They went close when Lilian Thuram served a cross from the right side, and Trezeguet's glancing header couldn't stay inside the left post (55'). France missed an even better opportunity soon after. Sylvain Wiltord served across from the right side, and Henry couldn't keep his header down, sailing it over an open goal. Then it was Yuri Djorkaeff's turn to have a go. His shot from 25 metres was stopped by Sylva, who had to scramble to punch the rebound over the by-line for a corner. Vieira then rose up to meet a corner from the right side and sent a powerful header on goal, forcing Sylva to make a nice diving save to his right (58').

Fußballbegriffe

31 May 2002, 19:31 Seoul – Tokyo FRANCE 0 : 1 SENEGAL by FIFAworldcup.com

FIFA World Cup' debutantes Senegal kicked off the 2002 edition of the world's greatest sporting event by shocking holders France 0 : 1 in Seoul, Korea through a first-half goal from Papa Bouba Diop. The French-speaking African side came out bravely, and behind the dangerous El Hadji Diouf, proved that they were not going to be intimidated. France increased their pressure but their finishing failed to break down a solid Senegalese defence. Considering that France hit the woodwork twice, Senegal coach Bruno Metsu could not be more pleased about the end result of the long-awaited match. Never has the loss of Zidane's magic wand been felt more cruelly by the French. Senegal did not seem intimidated at all by the FIFA World Cup holders in the early going, and El Hadji Diouf created the first chance for Senegal when he beat Marcel Desailly to the by-line and laid the ball back for Khalilou Fadiga, whose low shot rolled right to France goalkeeper Fabien Barthez. Shortly after France had a half-chance, when a Barthez punt was knocked on to a streaking David Trezeguet behind the Senegal defence. But the Juventus striker's first touch let him down. Trezeguet's touch failed him again minutes later as he struck the right post with a powerful, curling shot from the top of the penalty area (24'). Senegal shocked France with the tournament's first goal at the half-hour mark, just when it seemed that the French offside trap had Diouf in fits. The Senegalese striker rounded Frank Leboeuf and laid a dangerous ball across the face of the goal. It took deflections from Emmanuel Petit and Barthez before sitting up wonderfully for Papa Bouba Diop, who easily knocked the ball into the open net from his backside (0:1, 30'). After giving up the goal, France finished out the half by creating a flurry of dangerous situations, but they never managed to truly test Sylva. France opened the second half with a string of chances. They went close when Lilian Thuram served a cross from the right side, and Trezeguet's glancing header couldn't stay inside the left post (55'). France missed an even better opportunity soon after. Sylvain Wiltord served across from the right side, and Henry couldn't keep his header down, sailing it over an open goal. Then it was Youri Djorkaeff's turn to have a go. His shot from 25 metres was stopped by Sylva, who had to scramble to punch the rebound over the by-line for a corner. Vieira then rose up to meet a corner from the right side and sent a powerful header on goal, forcing Sylva to make a nice diving save to his right (58').

SProUt (DFKI): Grammatik für IE

- Gazetteer
- Morphologie
- Typenhierarchie
- Grammatiken für Namenserkennung
- Grammatiken für die Fußballdomäne
- Multilingual

RMRS-Output von SProUt

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---------------------|--|------------------|--|--|--|--|-----|-----------|--|--|--|------|-----------|--|--|--|------|--------------------|--|--|--|--|---------------------|--|--|--|--|-----|-----------|--|--|--|------|-----------|--|--|--|------|-------------|--|--|--|------|-----------|--|--|--|--|-----------------|--|--|--|--|-----|------------|--|--|--|------|------------|--|--|--|------|-----------|--|--|--|------|-----------|--|--|--|---|------------------|--|--|--|--|-----|------------|--|--|--|------|------------|--|--|--|------|-----------|--|--|--|------|-----------|--|--|--|--|-----------------|--|--|--|--|-----|------------|--|--|--|------|------------|--|--|--|------|-------------|--|--|--|------|-----------|--|--|--|
| TEXT | 31 May 2002 , 19 : 31 Seoul - Tokyo FRANCE 0 : 1 SENEGAL by FIFAworldcup.com FIFA World Cup ' debutantes Senegal kicked off the 2002 world'sgreatest sporting event by shocking holders France 0 : 1 in Seoul , Korea through a first-half goal from Papa Bouba Diop. The French came out bravely , and behind the dangerous ElHadji Diouf , proved that they were not going to be intimidated . | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| TOP | h0 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| RELS | <table border="0"> <tr> <td><i>point-rel</i></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>LBL</td> <td><i>h0</i></td> <td></td> <td></td> <td></td> </tr> <tr> <td>ARG0</td> <td><i>x0</i></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CARG</td> <td><i>31 May 2002</i></td> <td></td> <td></td> <td></td> </tr> </table> | <i>point-rel</i> | | | | | LBL | <i>h0</i> | | | | ARG0 | <i>x0</i> | | | | CARG | <i>31 May 2002</i> | | | | <table border="0"> <tr> <td><i>muc-type-rel</i></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>LBL</td> <td><i>h1</i></td> <td></td> <td></td> <td></td> </tr> <tr> <td>ARG0</td> <td><i>x1</i></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CARG</td> <td><i>date</i></td> <td></td> <td></td> <td></td> </tr> <tr> <td>ARG1</td> <td><i>x0</i></td> <td></td> <td></td> <td></td> </tr> </table> | <i>muc-type-rel</i> | | | | | LBL | <i>h1</i> | | | | ARG0 | <i>x1</i> | | | | CARG | <i>date</i> | | | | ARG1 | <i>x0</i> | | | | <table border="0"> <tr> <td><i>dofm-rel</i></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>LBL</td> <td><i>h17</i></td> <td></td> <td></td> <td></td> </tr> <tr> <td>ARG0</td> <td><i>x17</i></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CARG</td> <td><i>31</i></td> <td></td> <td></td> <td></td> </tr> <tr> <td>ARG1</td> <td><i>x0</i></td> <td></td> <td></td> <td></td> </tr> </table> | <i>dofm-rel</i> | | | | | LBL | <i>h17</i> | | | | ARG0 | <i>x17</i> | | | | CARG | <i>31</i> | | | | ARG1 | <i>x0</i> | | | | <table border="0"> <tr> <td><i>month-rel</i></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>LBL</td> <td><i>h18</i></td> <td></td> <td></td> <td></td> </tr> <tr> <td>ARG0</td> <td><i>x18</i></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CARG</td> <td><i>05</i></td> <td></td> <td></td> <td></td> </tr> <tr> <td>ARG1</td> <td><i>x0</i></td> <td></td> <td></td> <td></td> </tr> </table> | <i>month-rel</i> | | | | | LBL | <i>h18</i> | | | | ARG0 | <i>x18</i> | | | | CARG | <i>05</i> | | | | ARG1 | <i>x0</i> | | | | <table border="0"> <tr> <td><i>year-rel</i></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>LBL</td> <td><i>h19</i></td> <td></td> <td></td> <td></td> </tr> <tr> <td>ARG0</td> <td><i>x19</i></td> <td></td> <td></td> <td></td> </tr> <tr> <td>CARG</td> <td><i>2002</i></td> <td></td> <td></td> <td></td> </tr> <tr> <td>ARG1</td> <td><i>x0</i></td> <td></td> <td></td> <td></td> </tr> </table> | <i>year-rel</i> | | | | | LBL | <i>h19</i> | | | | ARG0 | <i>x19</i> | | | | CARG | <i>2002</i> | | | | ARG1 | <i>x0</i> | | | |
| <i>point-rel</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| LBL | <i>h0</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ARG0 | <i>x0</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CARG | <i>31 May 2002</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>muc-type-rel</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| LBL | <i>h1</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ARG0 | <i>x1</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CARG | <i>date</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ARG1 | <i>x0</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>dofm-rel</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| LBL | <i>h17</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ARG0 | <i>x17</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CARG | <i>31</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ARG1 | <i>x0</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>month-rel</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| LBL | <i>h18</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ARG0 | <i>x18</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CARG | <i>05</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ARG1 | <i>x0</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| <i>year-rel</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| LBL | <i>h19</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ARG0 | <i>x19</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| CARG | <i>2002</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ARG1 | <i>x0</i> | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| HCONS | { } | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| ING | { } | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

XML output

```
<F name="OUT">  
  <FS type="game_result">  
</FS>  
<F name="RESULT">  
  <FS coref="3" type=""0 : 1""/>  
</FS>  
</F>  
</F>
```

Vokabular der IE

- *Entitäten*: Interessante Objekte, wie Personen oder Organisationen
- *Attribute*: Eigenschaften der Objekte, wie z.B. der Name oder das Alter
- *Relationen*: Beziehungen zwischen Entitäten, wie z.B. zwischen einer Person und der Organisation
- *Ereignisse*: Aktivitäten, die die Entitäten betreffen, wie z.B. ein Unfall oder die Einführung eines neuen Produkts

Präzision der Extraktion, geschätzt

| Informationstyp | Präzision |
|-----------------|-----------|
| Entitäten | 90-98% |
| Attribute | ~80% |
| Relationen | 60-70% |
| Ereignisse | 50-60% |

Anwendungen der IE, z.B.

- Automatische Zusammenfassungen
- Automatische Fragebeantwortung
- Automatischer Aufbau von wissensbasierten Datenbanken
- Hyper-Linking
- Sentimentanalyse
- Trendanalyse
- Semantische Suche
- Visualisierung von Information

Warum ist IE schwierig?

- **Verschiedene Sprachen**
 - Morphologie ist einfach für das Englische, aber kompliziert für Deutsch und Türkisch
 - Segmentierung in Wörter ist einfach für Deutsch, aber kompliziert für Chinesisch und Japanisch
- **Verschiedene Texttypen**
 - Zeitung
 - Emails
 - Internet-Foren
 - Technische Dokumentation
- **Verschiedene Formate**
 - Tabellen
 - Fließtext
 - Kurzmeldungen

Warum ist IE schwierig?

- Ambiguitäten:
 - „Bank“ (Datenbank, Sitzbank, Geldinstitut), „Maus“
 - „Sieben“: „Der Wert ist: Sieben“, „Hier ist ein Stapel von Sieben“, „Vergessen Sie nicht das Sieben“
 - → Vokabular ist immer domänenspezifisch!
- Eigennamen
 - Können nicht alle aufgezählt werden
 - Können auch nicht alle trainiert werden
 - Es entstehen immer neue Eigennamen
 - Personen, Produkte, Orte, Firmen, Events
- komplexe Sätze
 - Verben und ihre Argumente – Events und ihre Akteure

Informationsextraktion auf Lageberichten

- Idee:

Analyse von Lageberichten für mittelgroße und große Kapitalgesellschaften in Deutschland

- Arbeiten von Christopher Almeling, Wolfram Bartussek und Melanie Siegel

- Wir bitten um sehr kritische Aufmerksamkeit und Diskussion!

Bundesanzeiger x Bundesanzeiger x +

https://www.bundesanzeiger.de/ebanzwww/wexssvlet?session.sessionid=374c034a00c70a8c6a94775bd697a524&g Suchen

Meistbesucht

Gefunden in:

Alle Bereiche (24683)

- » Amtlicher Teil (22)
- » Gerichtlicher Teil (9)
- » Gesellschaftsbe-
kannntmachungen (20032)
- » Kapitalmarkt (128)
- » Rechnungslegung/
Finanzberichte (4428)
- » Verschiedene
Bekanntmachungen (64)

Weiter eingrenzen:

In **Alle Bereiche** stehen Ihnen folgende Suchmöglichkeiten zur Verfügung

Veröffentlichungsdatum

von:

Tag ▼ Monat ▼ Jahr ▼

bis:

Tag ▼ Monat ▼ Jahr ▼

| Name ▼▲ | Bereich | Information ▼▲ | V.-Datum ▼▲ | Relevanz |
|--|------------------------------------|---|-------------------|----------|
| Kreissparkasse Ravensburg Ravensburg | Rechnungslegung/ Finanzberichte | » Jahresabschluss zum Geschäftsjahr vom 01.01.2014 bis zum 31.12.2014 | 04.08.2015 | ■■■■■ |
| ThyssenKrupp AG Duisburg und Essen | Rechnungslegung/ Finanzberichte | » Jahresabschluss zum Geschäftsjahr vom 01.10.2013 bis zum 30.09.2014 | 25.02.2015 | ■■■■■ |
| Peter Schneider Gebäudedienstleistungen GmbH & Co. KG Hannover | Rechnungslegung/ Finanzberichte | » Jahresabschluss zum Geschäftsjahr vom 01.01.2013 bis zum 31.12.2013 | 18.02.2015 | ■■■■■ |
| KAEFER Immobilien GmbH & Co.KG Bremen | Rechnungslegung/ Finanzberichte | » §§ 264 Abs. 3, 264b HGB zum Geschäftsjahr vom 01.01.2013 bis zum 31.12.2013 KAEFER Isoliertechnik GmbH & Co. Kommanditgesellschaft (Bremen) | 24.07.2014 | ■■■■■ |
| KAEFER Isoliertechnik GmbH & Co. Kommanditgesellschaft (Bremen) | Rechnungslegung/ Finanzberichte | » §§ 264 Abs. 3, 264b HGB zum Geschäftsjahr vom 01.01.2013 bis zum 31.12.2013 KAEFER Isoliertechnik GmbH & Co. Kommanditgesellschaft (Bremen) | 24.07.2014 | ■■■■■ |
| Caritas-Seniorenheime Betriebsführungs- und Trägerschaft GmbH Warendorf | Rechnungslegung/ Finanzberichte | » Jahresabschluss zum Geschäftsjahr vom 01.01.2012 bis zum 31.12.2012 Berichtigt am 29.01.2014 | 17.01.2014 | ■■■■■ |
| Institut für Angewandte Photovoltaik GmbH Gelsenkirchen | Rechnungslegung/ Finanzberichte | » Jahresabschluss zum Geschäftsjahr vom 19.12.2011 bis zum 31.12.2011 | 16.01.2014 | ■■■■■ |
| GELITA AG Eberbach | Rechnungslegung/ Finanzberichte | » Konzernabschluss zum Geschäftsjahr vom 01.01.2012 bis zum 31.12.2012 | 10.01.2014 | ■■■■■ |
| Caritas - Altenhilfe Dortmund GmbH Dortmund | Rechnungslegung/ Finanzberichte | » Jahresabschluss zum Geschäftsjahr vom 01.01.2012 bis zum 31.12.2012 | 26.11.2013 | ■■■■■ |
| Siemens Audiologische Technik GmbH Erlangen | Rechnungslegung/ Finanzberichte | » Jahresabschluss zum Geschäftsjahr vom 01.10.2011 bis zum 30.09.2012 | 14.08.2013 | ■■■■■ |

Die Geschäftsführung

Hansrüdiger Fritz

Martin Meyer-Luu

Lagebericht für das Geschäftsjahr 2011

der Nordwestbahn GmbH, Osnabrück

A. Wirtschaftliche Rahmenbedingungen

Der wirtschaftliche Aufschwung in Deutschland setzte sich im Jahr 2011 fort, sodass im Jahresverlauf das Bruttoinlandsprodukt das Vorkrisenniveau wieder erreicht hat.¹ Die gesamtwirtschaftliche Leistung in den ersten drei Quartalen des Jahres 2011 stieg dabei pro Quartal real um durchschnittlich 0,7%.² Auch im Vorjahresvergleich stieg die Wirtschaftsleistung deutlich an: Das preisbereinigte Bruttoinlandsprodukt lag im ersten Quartal 2011 um 5,0%, im zweiten Quartal 2011 um 3,0% und im dritten Quartal um 2,5% höher als im Vorjahreszeitraum.³

Die Abschwächung des wirtschaftlichen Aufschwungs zum Jahresende 2011 ist maßgeblich auf die Abkühlung des weltwirtschaftlichen Umfelds zurückzuführen, das bis dahin über die Exportnachfrage die deutsche Konjunktur stützte. Die Grundlage für die Ausweitung des Bruttoinlandsproduktes im zweiten Halbjahr 2011 bildete deshalb die Binnennachfrage. Im dritten Quartal wurde im Inland insbesondere mehr konsumiert, aber auch mehr investiert: Sowohl die privaten Haushalte mit einem Plus von 0,8% als auch der Staat mit einem Plus von 0,6% gaben deutlich mehr für Konsum aus als noch im zweiten Quartal 2011.⁴

Im Zuge der konjunkturellen Aufwärtsbewegung hat sich auch die Arbeitsmarktlage weiter verbessert: Die Arbeitslosenquote reduzierte sich im November 2011 auf 6,4%. Im Vergleich zum Vorjahr nahm die Zahl der arbeitslosen Menschen um 214.000 ab, wobei der Zuwachs an sozialversicherungspflichtiger Beschäftigung maßgeblich für den Rückgang ist. Damit erhöhen sich tendenziell die Chancen, durch Aufnahme einer Beschäftigung die Arbeitslosigkeit zu beenden. Neben den konjunkturellen Gründen haben der Strukturwandel und die Reformen am Arbeitsmarkt diese Chancen deutlich verbessert.⁵

Insgesamt zeigte sich die deutsche Wirtschaft ausgesprochen robust: Die positive Entwicklung auf dem Arbeitsmarkt und bei den Einkommen stützt den privaten Konsum und stabilisiert die binnenwirtschaftliche Entwicklung. Zwar wird sich der wirtschaftliche Aufschwung im Jahr 2012 voraussichtlich abschwächen, jedoch bedeutet dies nach der ausgeprägten Aufholphase eine Rückkehr zur Normalität.

¹ Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung, Jahresgutachten 2011/ 2012

² BMWi, Die wirtschaftliche Lage in Deutschland im Dezember 2011, Pressemitteilung vom 9. Dezember 2011

³ Destatis, Ausführliche Ergebnisse zur Wirtschaftsleistung im 3. Quartal 2011, Pressemitteilung Nr.435 vom 24. November 2011

⁴ Siehe ebenda.

⁵ Bundesagentur für Arbeit, Der Arbeitsmarkt im November: Anhaltend positive Entwicklung, Presse Info 072 vom 30. November 2011

Fragestellungen für Finanzanalysten und Wirtschaftswissenschaftler

- Wie hat sich die wirtschaftliche Lage im abgelaufenen Geschäftsjahr entwickelt?
- Welche Entwicklungen sind in der Zukunft zu erwarten?
- Welche Wirkungen hatte das unternehmerische Handeln auf Umwelt und Gesellschaft?
- Welche Maßnahmen hat das Unternehmen ergriffen?

Lageberichte aus der Sicht des Wirtschaftswissenschaftlers

- Große Textmengen, die gelesen werden müssen
- Information steht an unterschiedlichen Stellen in unterschiedlichen Lageberichten
- Viel Zeit geht dabei verloren, die Analysedaten in den Texten zu finden

Lageberichte aus der Sicht der Sprachtechnologin

- unstrukturierte Daten: Text
- Variation in der Ausdrucksweise:
 - derselbe Inhalt kann in verschiedenen Sätzen ausgedrückt werden
 - Tabellen haben in jedem Geschäftsbericht andere Spalten und Zeilen
 - unterschiedliche Überschriften
 - uneinheitliche Terminologie
- notwendig ist eine Technologie, die auf die Semantik der Texte Bezug nehmen kann

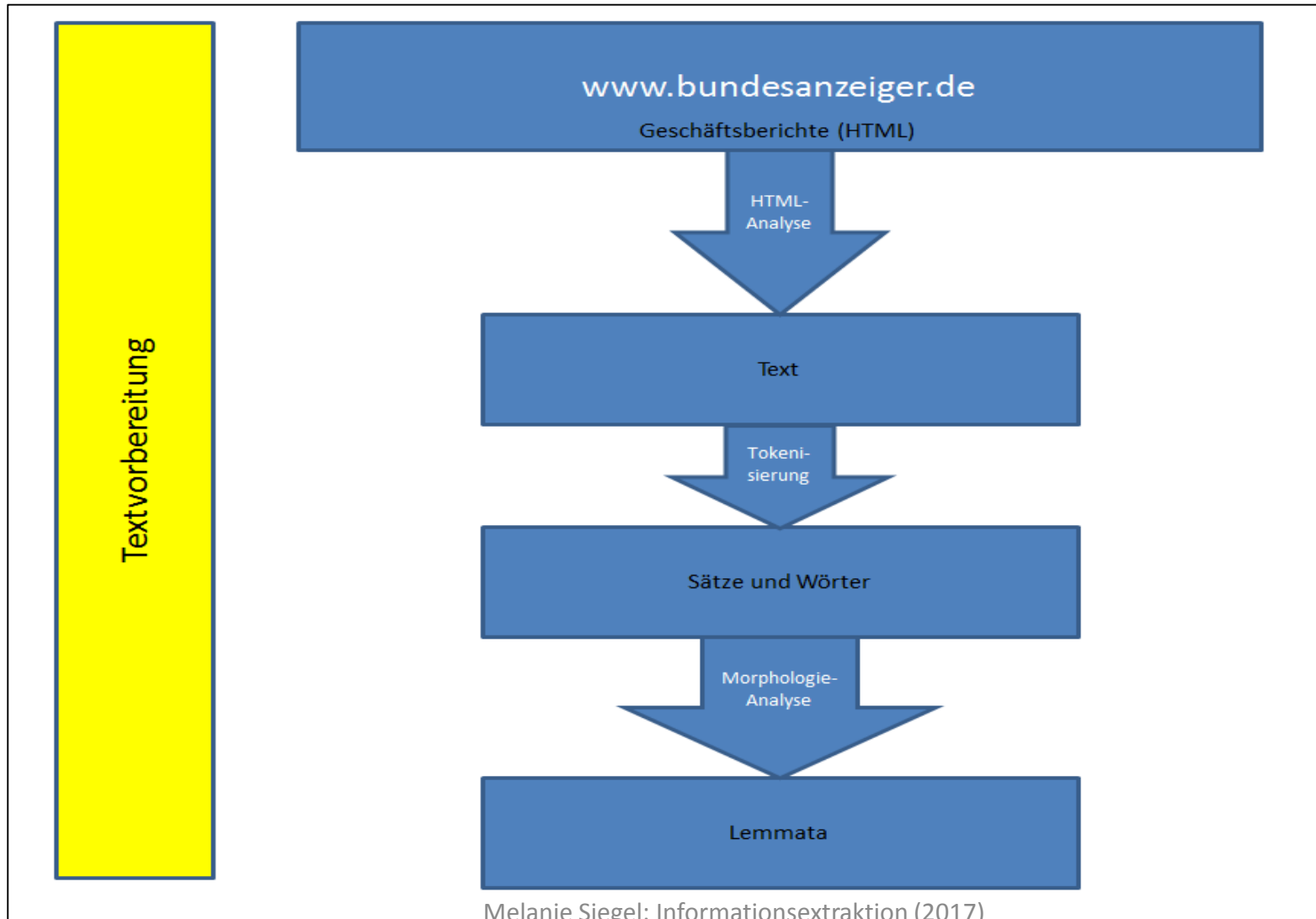
Fragestellung

Welche Technologien der Informationsextraktion können auf Geschäfts- und Lageberichte angewendet werden, um die Analyse der Finanzexperten zu unterstützen?

Analysebereiche

1. Analyse von wirtschaftlichen Prognosen
2. Analyse von Informationen über die Auswirkungen unternehmerischen Handelns auf Umwelt und Gesellschaft (CSR)

Vorbereitung der Texte und Analysen mit NLP-Methoden



Analyse von wirtschaftlichen Prognosen mit Pattern-Regeln

Wirtschaftliche Prognosen

- „Infolgedessen rechnen wir mit einer Steigerung des Ergebnisses je Aktie auf einen Wert zwischen 4,25 € und 4,40 €.“
- „Steigende Arbeitslosigkeit, strikte Sparmaßnahmen, geringe Lohnsteigerungen und niedrigere Konsumausgaben werden die Entwicklung des Sporteinzelhandels voraussichtlich negativ beeinflussen.“

XML-Format für die strukturierte Darstellung von Prognosen

```
<ANALYSIS>  
  <TEXT> Satz, in dem die Information steht </TEXT>  
  <PROGNOSE>  
    <ABOUT> Art der Prognose </ABOUT>  
    <ORGANIZATION> Firma oder Organisation </ORGANIZATION>  
    <MARKET> Markt, also z.B. der asiatische Markt </MARKET>  
    <DIVISON> Abteilung der Firma oder Organisation </DIVISON>  
    <PRODUCT> Produkt </PRODUCT>  
    <VALUE> Wert </VALUE>  
  </PROGNOSE>  
</ANALYSIS>
```

Pattern-Regeln für Named Entity Recognition

- Ziel: Felder ORGANIZATION, DIVISION, PRODUCT, MARKET und VALUE füllen
- Implementierung:
 - Python
 - Reguläre Ausdrücke
 - basierend auf Textblob POS Tags
 - Gazetteer-Listen mit Hinweisen auf Namen, z.B. „GmbH“, „AG“

Pattern-Regeln für Prognosen

- Ziel: Feld ABOUT
- Python, reguläre Ausdrücke

- z.B.:

```
steigerung_pattern =  
re.compile(r,  
(voraussichtlich|erwartet|rechnen|erwarten|prognostizieren|gehen).* .*  
  (Umsatz|Bruttomarge) .* (steigen|ansteigen)')
```

Beispielergebnisse Prognosen

```
1 <ANALYSIS>
2   <TEXT>Infolgedessen rechnen wir mit einer Steigerung des Ergebnisses je Aktie auf einen Wert zwischen 4,25 € und 4,40 €.</TEXT>
3   <PROGNOSE>
4     <ABOUT>Aktiensteigerung_auf</ABOUT>
5     <ORGANIZATION></ORGANIZATION>
6     <MARKET></MARKET>
7     <DIVISION></DIVISION>
8     <PRODUCT></PRODUCT>
9     <VALUE>einen Wert zwischen 4,25 € und 4,40 €</VALUE>
10  </PROGNOSE>
11 </ANALYSIS>
```

Beispielergebnisse Prognosen

```
12
13 <ANALYSIS>
14   <TEXT>Steigende Arbeitslosigkeit, strikte Sparmaßnahmen, geringe Lohnsteigerungen und niedrigere Konsumausgaben
15     werden die Entwicklung des Sporteinzelhandels voraussichtlich negativ beeinflussen.</TEXT>
16   <PROGNOSE>
17     <ABOUT>Umsatzbeeinträchtigung</ABOUT>
18     <ORGANIZATION></ORGANIZATION>
19     <MARKET></MARKET>
20     <DIVISION>Sporteinzelhandels</DIVISION>
21     <PRODUCT></PRODUCT>
22     <VALUE></VALUE>
23   </PROGNOSE>
24 </ANALYSIS>
25
```

Beispielergebnisse Prognosen

```
25
26 <ANALYSIS>
27   <TEXT>Für XYZ erwarten wir einen Anstieg des währungsbereinigten Umsatzes im mittleren einstelligen Bereich.</TEXT>
28   <PROGNOSE>
29     <ABOUT>Umsatzsteigerung_um</ABOUT>
30     <ORGANIZATION>XYZ</ORGANIZATION>
31     <MARKET></MARKET>
32     <DIVISON></DIVISON>
33     <PRODUCT></PRODUCT>
34     <VALUE>mittleren einstelligen Bereich</VALUE>
35   </PROGNOSE>
36 </ANALYSIS>
```


Analyse von CSR-Informationen in Lageberichten auf der Basis ontologischer Information

SemTAG

localhost:52434

Suchen

SemTAG Ontologien Geschäftsberichte Statistiken Über uns Kontakt Admin Log in

SemTAG

SemTAG ist der Prototyp eines Systems zur semantischen Analyse von Geschäftsberichten. Frei definierbare Aspekte, z.B. im Rahmen der sogenannten Corporate Social Responsibility (CSR) wie Ökologie oder Vielfalt und Chancengleichheit werden durch Begriffscluster unterlegt und als Themen in den Geschäftsberichten identifiziert und hervorgehoben. Das so gewonnene Zahlenmaterial über Worthäufigkeiten oder Wortnähen etc. wird statistisch ausgewertet und visualisiert.

Mit SemTAG wird das Ziel verfolgt, die Prüfung von Lageberichten speziell im Hinblick auf qualitative Aspekte zu unterstützen wie:

- Systeme und Prozesse
- Chancen und Risiken
- Planungen und Prognosen
- CSR-Informationen
- Sonstige (z.B. Kundentreue)
- Verweise

Geschäftsberichte

localhost:52434/reports/

Suchen

SemTAG Ontologien **Geschäftsberichte** Statistiken Über uns Kontakt Admin Log in

Geschäftsberichte

Import
Analyse
Visualisierung

© 2016 - SemTAG

localhost:52434/reports/

Erste Datenanalysen und Aufbau der Wissensbasis

- Ziel: Suche nach Schlüsselwörtern und Entdeckung der Möglichkeiten semantischer Technologien
 1. Automatische Terminologie-Extraktion
 2. Aufstellung von Schlüsselwörtern
 3. Kookkurrenz-Analyse
 4. Komposita-Analyse

Automatische Terminologie-Extraktion

- Nutzung der Software Acrolinx
- Ergebnis:
 - Liste von Fachbegriffen
 - z.B. Finanzkennzahl, Fälligkeitenstruktur , Compliance
 - Terminologie-Varianten
 - z.B. Finanzkennzahl / Finanz-Kennzahl, Fälligkeitenstruktur / Fälligkeitsstruktur, XETRA-Handel / Xetra[®]-Handel

Auswertung der Terminologielisten

- Aufstellung einer Liste von Schlüsselwörtern durch den Wirtschaftswissenschafts-Experten
- Zuordnung zu den Indikatoren der G4-Leitlinien zur Nachhaltigkeitsberichterstattung der Global Reporting Initiative

Global Reporting Initiative (2015)

Sprachtechnologische Analyse der Lageberichte mit Schlüsselwörtern

- Kookkurrenzanalyse

- Ziel: Welche Wörter treten häufig gemeinsam mit diesen Schlüsselwörtern auf?

- Python (NLTK)-Textmining-Verfahren

- z.B. „Compliance“:

[('Risiken', 32), ('Aufsichtsrat', 16), ('Vorstand', 16), ('Groß', 15), ('Konzerns', 14), ('Dr.', 13), ('Allianz', 12), ('Daimler', 12), ('Moderat', 11), ('AG', 11), ('Konzern', 11), ('Mitarbeiter', 10), ('Group', 9), ('Prüfungsausschuss', 8), ('Unternehmens', 8), ('Ausschuss', 8), ('Überwachung', 8), ('Wirksamkeit', 8), ('Compliance-Risiken', 7), ('Richtlinien', 7), ('Mitglieder', 7), ('Einhaltung', 7), ('Fragen', 7), ('Geschäftsentwicklung', 7), ('Anteilseignervertreter', 7), ('Risikomanagementsystem', 7), ('berichtet', 7), ('Officer', 7), ('Risikomanagement', 7), ('Chief', 7), ('Insurance', 7), ('Aufsichtsrats', 6), ('Kontrollen', 6), ('Rahmen', 6), ('Integrität', 6), ('Perlet', 6), ('Kontrollsystems', 6), ('Risikomanagements', 6), ('Compliance-Organisation', 6), ('Risikomanagementsystems', 6), ('Legal', 6), ('Risk', 6), ('Jahr', 6), ('Helmut', 6), ('Dauer', 5), ('Revisionssystem', 5), ('Entwicklungen', 5)]

Sprachtechnologische Analyse der Lageberichte mit Schlüsselwörtern

- Kompositaanalyse
- Ziel: Erweiterung der Schlüsselwörter-Listen
- z.B. „Umwelt“:
['Auto-Umwelt-Ranking', 'US-Umweltschutzbehörde', 'Umweltangelegenheiten', 'Umweltanstrengungen', 'Umweltaspekte', 'Umweltauswirkungen', 'Umweltbelastung', 'Umweltbereich', 'Umweltbestimmungen', 'Umweltbilanz', 'Umweltdaten', 'Umweltfreundlichkeit', 'Umweltleistung', 'Umweltleistungen', 'Umweltmanagement', 'Umweltmanagementsysteme', 'Umweltnormen', 'Umweltpraktiken', 'Umweltpreis', 'Umweltrichtlinien', 'Umweltrisiken', 'Umweltschonung', 'Umweltschutz', 'Umweltschutzmaßnahmen', 'Umweltschutzrisiken', 'Umweltstandards', 'Umweltstrategie', 'Umweltverantwortung', 'Umweltverfahren', 'Umweltverträglichkeit', 'Umweltwissenschaftler', 'Umweltzeichen', 'Umweltzertifikat', 'Umweltzonen']

Aufstellung einer Ontologie

- Ziel: Organisation der Indexterme für eine Inhaltsanalyse
- Ergebnis der Datenanalyse:
 - Liste von Schlüsselwörtern und ihrer Zuordnung zu den Indikatoren der Global Reporting Initiative
 - Indikatoren sind bereits hierarchisch organisiert in einer Taxonomie mit drei Ebenen

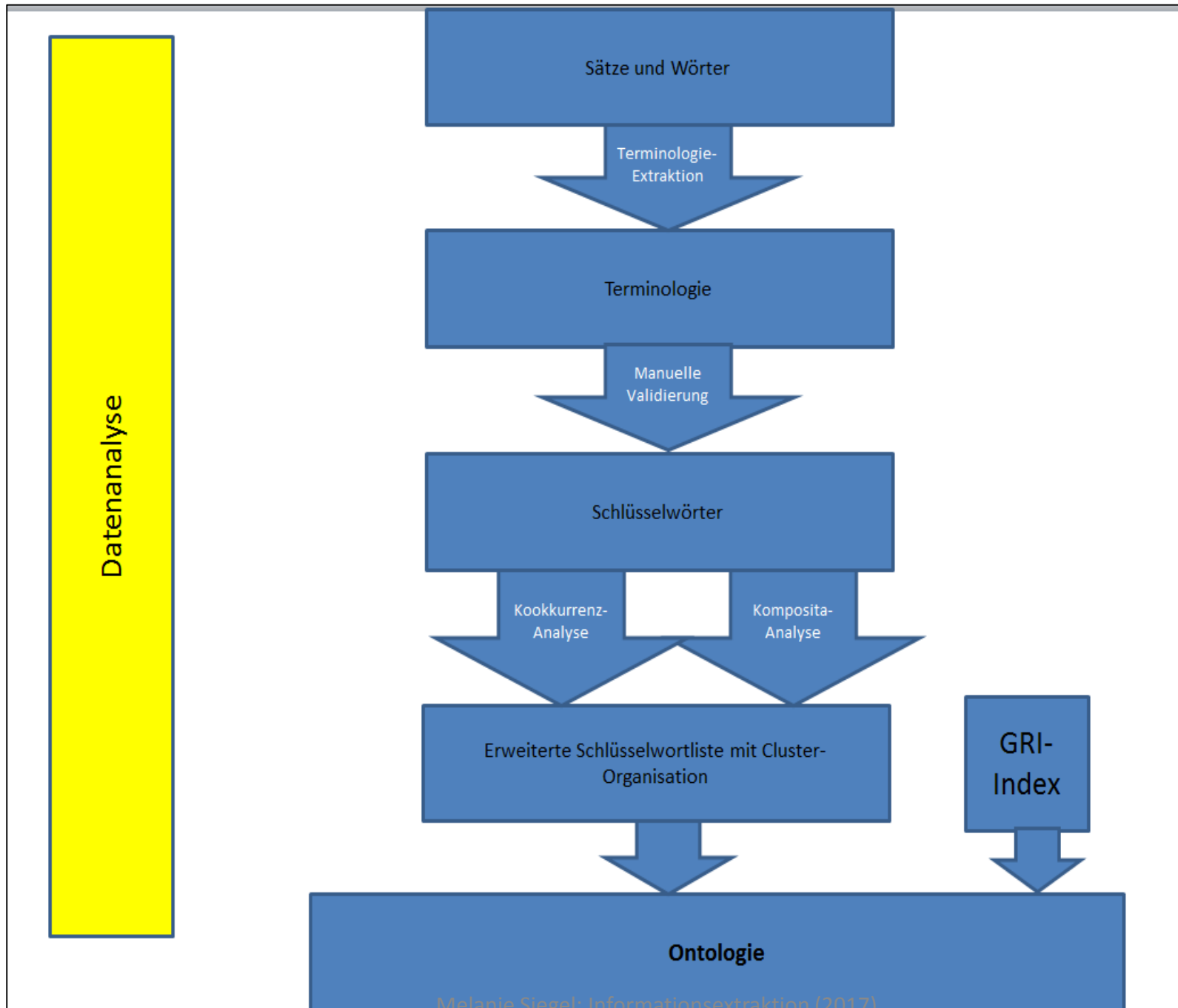
Taxonomie der GRI (Auszug)

| |
|--|
| Gesellschaft |
| <ul style="list-style-type: none">• Arbeitspraktiken und menschenwürdige Beschäftigung<ul style="list-style-type: none">○ Beschäftigung○ Arbeitgeber-Arbeitnehmer-Verhältnis○ Arbeitssicherheit und Gesundheitsschutz○ Aus- und Weiterbildung○ Vielfalt und Chancengleichheit○ Gleicher Lohn für Frauen und Männer○ Bewertung der Lieferanten hinsichtlich Arbeitspraktiken○ Beschwerdemechanismen hinsichtlich Arbeitspraktiken |
| <ul style="list-style-type: none">• Menschenrechte<ul style="list-style-type: none">○ Investment○ Gleichbehandlung○ Vereinigungsfreiheit und Recht auf Kollektivverhandlungen○ Kinderarbeit○ Zwangs- oder Pflichtarbeit○ Sicherheitspraktiken○ Rechte der indigenen Bevölkerung○ Prüfung○ Bewertung der Lieferanten hinsichtlich Menschenrechten○ Beschwerdemechanismen hinsichtlich Menschenrechtsverletzungen |
| <ul style="list-style-type: none">• Gesellschaft<ul style="list-style-type: none">○ Lokale Gemeinschaften○ Korruptionsbekämpfung○ Politik○ Wettbewerbswidriges Verhalten○ Compliance○ Bewertung der Lieferanten hinsichtlich gesellschaftlicher Auswirkungen○ Beschwerdemechanismen hinsichtlich gesellschaftlicher Auswirkungen |

Analyse der Taxonomie

- multiple Vererbung:
 - z.B. „Compliance“ steht in „Produktverantwortung“, „Gesellschaft“ und „Ökologie“
- hyponymische und meronymische Relationen
- verschiedene Relationen der Schlüsselwörter

→ Organisation in einer Ontologie



The screenshot shows a web browser window with the following elements:

- Browser Tab:** Labeled "SemTAG".
- Address Bar:** Contains "localhost:60407/ontoedit/".
- Search Bar:** Labeled "Suchen".
- Navigation Menu:** Includes "SemTAG", "Ontologien", "Geschäftsberichte", "Statistiken", "Über uns", "Kontakt", "Admin", and "Log in".
- Main Content Area:**
 - Section Header:** "Ontologien" (highlighted in a light gray box).
 - Sub-sections:** "Import", "Bearbeitung", and "Visualisierung".
 - Footer:** "© 2016 - SemTAG".

Select terms to change | SemT... X +

localhost:60407/admin/ontoedit/terms/?q=I3

SemTAG Administration WELCOME, ADMIN VIEW SITE / DOCUMENTATION / CHANGE PASSWORD / LOG OUT

Home > Ontoedit > Terms

Select terms to change

ADD TERMS +

Q I3 Search 4 results (1053 total)

Action: ----- Go 0 of 4 selected

| <input type="checkbox"/> | ID | TERM | RELTYPE | RELATED TO | RELATED TO TERM |
|--------------------------|------|-----------------|------------|------------|-----------------|
| <input type="checkbox"/> | 1051 | I3_wirtschaft | belongs_to | 1053 | I4_public_value |
| <input type="checkbox"/> | 1050 | I3_sonstiges | belongs_to | 1053 | I4_public_value |
| <input type="checkbox"/> | 1046 | I3_oekologie | belongs_to | 1053 | I4_public_value |
| <input type="checkbox"/> | 1045 | I3_gesellschaft | belongs_to | 1053 | I4_public_value |

4 terms

FILTER

By ID

- All
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19

Grundlegende Statistiken über Wertschöpfungsinformation in Lageberichten

- Ziel: Messung der Nachhaltigkeitsberichterstattung in deutschen Lageberichten
- Fragestellungen:
 1. Welche Themen werden im Geschäftsbericht behandelt? Sind die Angaben vollständig?
 2. Was ist der Umfang der Berichterstattung über Umwelt, Gesellschaft und Wirtschaft in Lageberichten im Verhältnis zum Gesamttext?
 3. Wie ist die Entwicklung des Themengebiets in Geschäfts- und Lageberichten im Zeitablauf? Wie ist das Themengebiet im Branchenvergleich behandelt?

Statistiken

localhost:52434/statistics/

Suchen

SemTAG Ontologien Geschäftsberichte **Statistiken** Über uns Kontakt Admin Log in

Statistiken

- Systeme und Prozesse
- Chancen und Risiken
- Planungen und Prognosen
- CSR-Informationen
- Sonstige
- Verweise

© 2016 - SemTAG

localhost:52434/statistics/

Welche Themen werden im Lagebericht behandelt? Sind die Angaben vollständig?

- Automatische Indexierung mit der aufgestellten Ontologie:
 - Text wird auf Schlüsselwörter untersucht.
 - Schlüsselwörter werden GRI-Themenbereichen zugeordnet.
 - Sätze mit Themenbereichen werden markiert.

Ausschnitt aus der Themenannotation

| | A | B | C | D | E | F | G |
|-----|-----------------------------|------------------------|---|---|---|---|---|
| 1 | Thema | Zahl der Wörter | Text | | | | |
| 98 | ['beschaeftigung'] | 14 | Personal Die Mitarbeiterzahl der Merck KGaA lag zum 31. Dezember 2012 bei 9.203 Beschäftigten. | | | | |
| 99 | ['beschaeftigung'] | 13 | Die Anzahl der Mitarbeiter hat sich gegenüber dem Vorjahr leicht reduziert (2011: 9.296). | | | | |
| 100 | ['arbeitspraktien_und_me | 40 | Aufgrund der Neuausrichtung lokaler Funktionen in globale Einheiten zu Beginn des Jahres kam es in allen Bereichen zu Veränderungen der Mitarbeiterstruktur Der Fraue | | | | |
| 101 | ['vielfalt_und_chancenglei | 18 | Merck hat weiterhin das Ziel, den Frauenanteil zu erhöhen - insbesondere dort, wo dieser unterrepräsentiert ist, beispielsweise in Führungspositionen. | | | | |
| 102 | ['vielfalt_und_chancenglei | 21 | Derzeit liegt der Anteil von Frauen in Führungspositionen konzernweit bei 24 % (2011: 23 %), in der Merck KGaA bei 20 % (2011: 17 %). | | | | |
| 103 | ['arbeitspraktien_und_me | 41 | Um den Anteil von Frauen in Führungspositionen bis 2016 weltweit auf 25 % bis 30 % zu erhöhen, sind zusätzlich zu den bereits bestehenden Maßnahmen wie Mentoring, | | | | |
| 104 | [] | 15 | Für Darmstadt liegt der Fokus 2013 auf der Etablierung von Trainings zu Selbstmarketing und Führung. | | | | |
| 105 | ['beschaeftigung'] | 18 | Die Arbeit, der Einfallsreichtum und das unternehmerische Handeln unserer Mitarbeiter sind entscheidend, damit das Unternehmen langfristig erfolgreich ist. | | | | |
| 106 | ['beschaeftigung'] | 12 | Wir fordern von unseren Beschäftigten eine hohe Leistung und honorieren diese entsprechend. | | | | |
| 107 | ['beschaeftigung'] | 11 | Alle Mitarbeiter erhalten in Abhängigkeit vom Erfolg des Unternehmens einen Unternehmensbonus. | | | | |
| 108 | ['forschung'] | 23 | Die Zielwerte hierfür wurden für 2012 für alle Mitarbeiter über Merck Scorecards, die Kennzahlen aus den Bereichen Finanzen, Mitarbeiter und Innovation enthalten, defini | | | | |
| 109 | [] | 18 | Je nach organisatorischer Zuordnung des Mitarbeiters ist die Scorecard der Division, des Business Sektors oder der Merck-Gruppe bonusrelevant. | | | | |
| 110 | [] | 21 | Ein Unternehmensbonus kommt im April 2013 zur Auszahlung, sofern die Zielerreichung der jeweiligen Scorecard für das Jahr 2012 über 50 % liegt. | | | | |
| 111 | [] | 17 | Bei einer Zielerreichung der Scorecard von 100 %, beläuft sich der Unternehmensbonus im Tarifbereich auf 40 % eines Monatsentgelts. | | | | |
| 112 | [] | 23 | Bei den außertariflichen Angestellten inklusive den Führungskräften liegt der Unternehmensbonus je nach Funktion in der Spanne zwischen 39 % eines Monatsgehalts un | | | | |
| 113 | ['bezahlung'] | 19 | Zusätzlich gibt es im außertariflichen Bereich individuelle leistungsbezogene Vergütungskomponenten, die davon abhängen, wie der Einzelne seine Ziele erreicht hat. | | | | |
| 114 | ['arbeitspraktien_und_me | 21 | Insgesamt liegt die variable Vergütung (individueller Bonus und Unternehmensbonus) bei den außertariflichen Angestellten und Führungskräften zwischen 10 % und 45 % | | | | |
| 115 | ['bezahlung'] | 15 | Seit 2009 gewährt Merck berechtigten Führungskräften eine langfristige erfolgsabhängige Vergütungskomponente, den Merck Long-Term Incentive Plan. | | | | |
| 116 | [] | 37 | Die Höhe der Auszahlung für die Tranche 2010 und 2011 ist abhängig von dem Erreichen der beiden Verhältniskennzahlen "auf die Gesamterlöse bezogener Basis Free Ca | | | | |
| 117 | [] | 25 | Bei dem Plan sind die potenziellen Auszahlungen in der Zukunft bei hohem Zielerreichungsgrad nach oben begrenzt - bei einem zu geringen Zielerreichungsgrad erfolgt das | | | | |
| 118 | [] | 34 | Die Auszahlung für die Tranche 2012 richtet sich danach, wie sich die Merck-Aktie im Vergleich zur Entwicklung des DAX30 sowie wie sich die beiden Finanzkennzahlen Ne | | | | |
| 119 | ['beschaeftigung', 'energie | 75 | Die Merck KGaA verzeichnete eine Fluktuationsrate von (bereinigt um Austritte wegen Pensionierung, Tod, befristete Arbeitsverhältnisse sowie Wechsel zu Tochterfirme | | | | |
| 120 | [] | 34 | Die Treibhausgas-Emissionen werden seit 2008 nach dem GHG Protocol (GHG = Green House Gas) erfasst, einem anerkannten System, das von World Resources Institute (V | | | | |
| 121 | [] | 20 | Um unser Klimaziel bis zum Jahr 2020 zu erreichen, legen wir im Rahmen unseres Klimaschutzprogramms EDISON den Schwerpunkt auf ca. | | | | |

Was ist der Umfang der Berichterstattung über Umwelt, Gesellschaft und Wirtschaft in Lageberichten im Verhältnis zum Gesamttext?

| | |
|---|--------------|
| Zahl der Wörter im Text | 83092 |
| Zahl der Sätze im Text | 4086 |
| Zahl der Wertschöpfungswörter im Text | 1853 |
| Prozentsatz der Wertschöpfungswörter an allen Wörtern | 2.23 % |
| Zahl der Sätze mit Wörtern aus dem Themenbereich Wertschöpfung | 1262 |
| Prozentsatz der Sätze mit Wertschöpfungsinhalten an allen Sätzen | 30.89 % |
| Zahl der Gesellschaftswörter im Text | 1304 |
| Prozentsatz der Gesellschaftswörter an allen Wörtern | 1.57 % |
| Zahl der Sätze mit Wörtern aus dem Themenbereich Gesellschaft | 943 |
| Prozentsatz der Sätze mit Gesellschaftsinhalten an allen Sätzen | 23.08 % |
| Zahl der Wirtschaftswörter im Text | 440 |
| Prozentsatz der Wirtschaftswörter an allen Wörtern | 0.53 % |
| Zahl der Sätze mit Wörtern aus dem Themenbereich Wirtschaft | 363 |
| Prozentsatz der Sätze mit Wirtschaftsinhalten an allen Sätzen | 8.89 % |
| Zahl der Ökologiewörter im Text | 758 |
| Prozentsatz der Ökologiewörter an allen Wörtern | 0.91 % |
| Zahl der Sätze mit Wörtern aus dem Themenbereich Ökologie | 609 |
| Prozentsatz der Sätze mit Ökologie-Inhalten an allen Sätzen | 14.90 % |
| Zahl der sonstigen Wörter im Text | 570 |
| Prozentsatz der sonstigen Wörter an allen Wörtern | 0.69 % |
| Zahl der Sätze mit Wörtern aus dem Themenbereich Sonstiges | 475 |
| Prozentsatz der Sätze mit Sonstiges-Inhalten an allen Sätzen | 11.63 % |

Wie ist die Entwicklung des Themengebiets in
Geschäfts- und Lageberichten im Zeitablauf?
Wie ist das Themengebiet im Branchenvergleich
behandelt?

- Vergleich des Textanteils zum Thema Wertschöpfung in den
Geschäftsberichten verschiedener Firmen
- Veränderung des Textanteils zum Thema Wertschöpfung in den
Geschäftsberichten über mehrere Jahre

Fazit

- Informationsextraktion geht über Text Mining hinaus und erfordert semantische Analysemethoden
- IE-Methoden können u.a. Wirtschaftswissenschaftler in ihren Analysen wirksam unterstützen
 - Extraktion von Information und Darstellung in strukturierter Form (XML, Tabelle, Datenbank,...)
 - Aufstellung einer Ontologie
 - Indexierung und statistische Auswertung der wirtschaftswissenschaftlichen Texte
- weitere Einsatzszenarien der Technologie sind denkbar – Ideen?

Über uns

localhost:60407/about

SemTAG Ontologien Geschäftsberichte Statistiken Über uns Kontakt Admin Log in

Über uns

Frei definierbare Aspekte, z.B. im Rahmen der sogenannten Corporate Social Responsibility (CSR) wie Ökologie oder Vielfalt und Chancengleichheit werden durch Begriffscluster unterlegt und als Themen in den Geschäftsberichten identifiziert und hervorgehoben. Das so gewonnene Zahlenmaterial z.B. über Worthäufigkeiten oder Wortnähren etc. wird statistisch ausgewertet und visualisiert.

SemTAG entsteht in Kooperation zwischen der

Hochschule Darmstadt, vertreten durch

Prof. Dr. Christopher Almeling, FB Wirtschaftswissenschaften, Rechnungswesen und Wirtschaftsprüfung

Prof. Dr. Melanie Siegel, FB Media, Computerlinguistik

und der

OntoPort UG (www.ontoport.de), Semantische Technologien, Data Science, Data Visualization

© 2016 - SemTAG

Kontakt

localhost:60407/contact

Suchen

SemTAG Ontologien Geschäftsberichte Statistiken Über uns **Kontakt** Admin Log in

Kontakt

Hochschule Darmstadt
Max-Planck-Str. 2
64807 Dieburg

OntoPort UG
Altkönigstr. 3
65843 Sulzbach/Taunus

Support: support@ontoport.de

© 2016 - SemTAG

localhost:60407/contact