

How to Boost Customer Relationship Management via Web Mining: Benefiting from the Glass Customer's Openness

Case Study: Overview

- *Company W* is active in the B2B sector and supplies craft businesses (e-commerce and traditional sales).
- Due to the highly competitive market, it is necessary to react quickly to relevant changes among clients.
- **Relevant categories:** Client's foundation year, number of employees and business area.
- Growing number of clients: manual maintenance of the CRM database is no longer possible.

→ Can Web Mining (WM) with Local Grammars (LGs) be used for **reliable data extraction** for CRM systems?

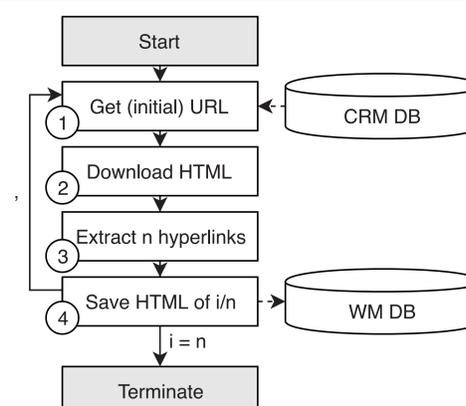
Examples (translated from German):

1. Our staff currently consists of **ten** employees
2. Since **1996** we are your painting company in Berlin
3. Our company was founded in the **early sixties**
4. Our team: **3** painters, **1** secretary and **me**
5. **1997:** The company celebrates its **120th anniversary** with seven employees

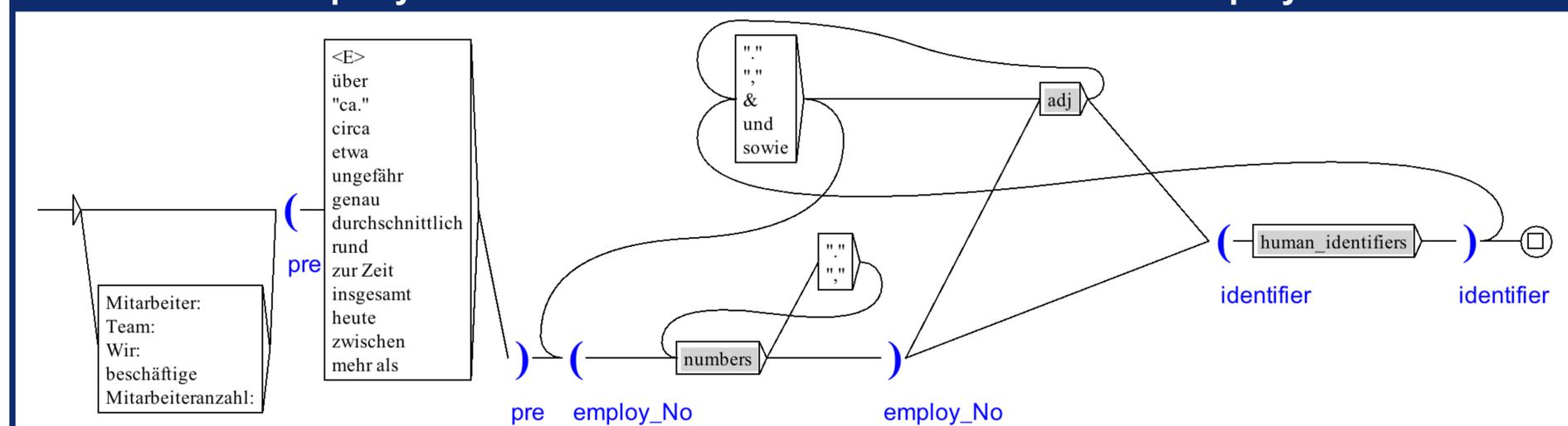
Web Mining System Architecture

Monitoring & Mining

- Iterative acquisition of websites up to a **depth of four subpages**
 - Use of filters and classification (binary) to find **relevant phrases**
 - **Max. number** of 150 pages per client, avg. **acquisition time** is 3.5 min.
 - Poor data quality, **poorly developed web pages** in this domain
- LGs can achieve a high recall on web content, but require extensive preprocessing



Exemplary Local Grammar: Extraction of the Number of Employees



Evaluation

- We designed a **manually annotated test corpus** composed of 1,000 sentences.
- Extraction of relevant categories using LGs: **Average F-Score** is 74% (avg. P. 87%, R. 64%)
- 83% of the websites could be assigned to a **business area** on basis of extracted keywords.
- 31% of all processed websites provide information on **foundation** year of the respective company.
- Only from 12% of the websites could the number of employees be extracted.

Conclusion

- LGs can be used for Web Mining CRM systems.
- The challenge for LGs is the **surprisingly poor quality** of the NL texts and the lack of context.
- **Poor text quality** can be partly compensated by weakening the LGs (damages precision), missing context cannot be compensated.
- In future work, we will create more LGs to acquire **further information** for the B2B CRM.
- Moreover, we will improve the reliability of the current LGs by **improving linguistic resources**.