Coping with Evolving Knowledge
Dynamic Domain Ontologies for Web Intelligence

5 September 2012

albert.weichselbraun@htwchur.ch

Member of the University of Applied Sciences Eastern Switzerland (FHO)
Agenda

Part I: Ontologies for Web Intelligence
- Web Intelligence
- NOAA Media Watch
- Applications of ontologies for Web Intelligence tasks

Part II: Dynamic Lightweight Domain Ontologies
- Ontology learning from heterogeneous sources
- Coping with change
- Outlook and conclusions
Part I
Ontologies for Web Intelligence

Powered by webLyzard technology

Member of the University of Applied Sciences Eastern Switzerland (FHO)
Web Intelligence

- **Business Intelligence** has been used as an umbrella term to describe concepts and methods for improving business decision making by using fact-based support systems (Chen 2010).

- Luhn (1958): “A business intelligence system”

- Structured data sources (databases, warehouses, …)

- Case studies
  - Continental Airlines (Anderson-Lehman et al. 2004)
  - Parkway (Negash and Gray 2008)
  - 1-800-CONTACTS (Watson and Wixom 2007)
NOAA collects 80 TB of scientific data per day, with a 10-fold increase expected by 2020. Annual IT Budget: USD 1 Billion
July global temperatures fourth highest on record; Arctic sea ice is second lowest July extent on record
According to NOAA NCDC scientists, the globally-averaged temperature for July 2012 marked the fourth warmest July since record keeping began in 1880. It also marked the 36th consecutive July and 329th consecutive month with a global temperature above the 20th century average.

Updated Atlantic hurricane outlook calls for near- or above-normal season
This year’s Atlantic hurricane season got off to a busy start, with 6 named storms to date, and may have a busy second half, according to the updated hurricane season outlook issued today by NOAA’s Climate Prediction Center, a division of the National Weather Service.

July 2012 marked the hottest month on record for the contiguous United States
Drought expands to cover nearly 63 percent of the Lower 48; wildfires consume two
Data Acquisition & Contextualization
Warming causes more extreme shifts

Warming causes more extreme shifts of the Southern Hemisphere's largest rain band. Occasionally, the rain band moves northwards towards the Equator by 1000 kilometres, inducing extreme climate events. In turn, and in spite of disagreement about the future of El Niño events, this warming leads to the increased frequency of extreme excursions of the rain band. During moderate El Niño events with warming in the equatorial eastern Pacific, the rain band moves north-eastward by 300 kilometres.
Tag Cloud on Climate Change

News Media

Social Media

Fortune 1000
Analytical Questions & Applications

- How can we detect, describe and potentially predict patterns in online coverage (external events versus resonance patterns)?
  
  MARKET RESEARCH, TREND SCOUTING

- How widespread is content redundancy, and what influences content replication within and across networks (e.g. network and community characteristics)? Who are the most influential opinion leaders driving this process?
  
  OPINION LEADER ANALYSIS, IMPACT MEASURES, VIRAL MARKETING STRATEGIES

- How do macroscopic information flows shape public opinion? What are appropriate methods to measure and model the extent, dynamics and latency of this process?
  
  FINANCIAL MARKETS, POLITICAL CAMPAIGNS, SCIENCE COMMUNICATION
Ontologies for Sentiment Analysis

- Problem: Ambiguities, Context
- Data Sources: ConceptNet, Wordnet

**Workflow:**

1. Identify ambiguous terms based on a training corpus
2. Determine the context for ambiguous terms → contextualized sentiment dictionary
3. Use the contextualized sentiment dictionary for ambiguous terms
Ontologies for Geotagging

- Problem: Ambiguous Locations, Focus
- Data Sources: GeoNames

**Workflow:**

1. Build a data model based on the GeoNames ontology and the corresponding instance data.
   → nearbyFeatures, parentFeature, parentCountry, population

2. Use this context information for disambiguation and focus selection.
Further Application Areas

- Navigation
- Search query expansion
- Pre-structuring of the semantic map
- Named entity detection
Part II
Dynamic Domain Ontologies

Powered by webLyzard technology

Member of the University of Applied Sciences Eastern Switzerland (FHO)
Ontology Learning

- ontology learning framework for lightweight domain ontologies (Hendler 2009, Alani et al. 2008)
- based on a seed ontology and domain documents
  - extract relevant terms
  - integrate them into the ontology
- benefits of integrating social sources
  - potential of providing background knowledge
  - contain the latest terminology (Angeletou et al. 2007)
  - (evolve at much a higher pace than domain documents)
Architecture - Overview
# Suggested Terms

<table>
<thead>
<tr>
<th>Unstructured data sources</th>
<th>Social delicious</th>
<th>Social flickr</th>
</tr>
</thead>
<tbody>
<tr>
<td>targets</td>
<td>animalcare</td>
<td>architecture</td>
</tr>
<tr>
<td>building</td>
<td>architects</td>
<td>art</td>
</tr>
<tr>
<td>coal</td>
<td>atmosphere</td>
<td>auckland</td>
</tr>
<tr>
<td>levels</td>
<td>award</td>
<td>beach</td>
</tr>
<tr>
<td>climate change policy</td>
<td>carbonfoodprint</td>
<td>bicycle</td>
</tr>
<tr>
<td>pact</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>reduce greenhouse gas</td>
<td>technorati</td>
<td>twitter</td>
</tr>
<tr>
<td>pollution</td>
<td>agile</td>
<td>aces</td>
</tr>
<tr>
<td>firm</td>
<td>apple</td>
<td>afgan</td>
</tr>
<tr>
<td>carbon dioxide emissions</td>
<td>art.</td>
<td>afganistan</td>
</tr>
<tr>
<td>etcs</td>
<td>automotice</td>
<td>africa</td>
</tr>
</tbody>
</table>
## Suggested Terms

<table>
<thead>
<tr>
<th>Seed Concept (Cₜ)</th>
<th>Evidence-Source (e)</th>
<th>Candidate Concept (C_c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>climate_change</td>
<td>oe:coOccurs</td>
<td>“CO2”</td>
</tr>
<tr>
<td>_:1</td>
<td>rdf:subject</td>
<td>climate_change</td>
</tr>
<tr>
<td>_:1</td>
<td>rdf:predicate</td>
<td>oe:coOccurs</td>
</tr>
<tr>
<td>_:1</td>
<td>rdf:object</td>
<td>“CO2”</td>
</tr>
<tr>
<td>_:1</td>
<td>rdf:type</td>
<td>rdf:Statement</td>
</tr>
<tr>
<td>_:1</td>
<td>oe:significance</td>
<td>“3.20”</td>
</tr>
<tr>
<td>climate_change</td>
<td>ow:twitter</td>
<td>“CO2”</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>climate_change</td>
<td>wn:hyponym</td>
<td>temperature_change</td>
</tr>
</tbody>
</table>
Concept Selection and Positioning
Concept Selection and Positioning
Concept Selection and Positioning
Concept Selection and Positioning

[Image of a network diagram with nodes such as energy, india, policy, climate, co2, climate_change, china, malcom, copenhagen, fossil_fuel, coal, kyoto, methan, greenhouse_gas, cfc, and labels indicating connections and weights between them.]
Concept Selection and Positioning
Domains Evolve

- **Static Approach**
  - Fixed transformation functions
    (evidence sources $\rightarrow$ spreading activation weights)
  - Determined by domain experts

- **Problems**
  - Domains evolve
  - Importance and reliability of evidence sources changes

- “Gold Standard”
  - Users
  - Games with a Purpose
Learning Evolving Ontologies
Relation Strength Matrix

- A formalized representation of the ontology
- Records the relation strength ($rs_{ij}$) for any relation $r_{ij}$ between seed concept “i” and candidate concept “j”.

$$RSM = \begin{bmatrix}
rs_{r_{11}t_{1}} & rs_{r_{11}t_{2}} & \cdots & rs_{r_{11}t_{n}} \\
rs_{r_{12}t_{1}} & rs_{r_{12}t_{2}} & \cdots & rs_{r_{12}t_{n}} \\
\vdots & \vdots & \ddots & \vdots \\
rs_{r_{21}t_{1}} & rs_{r_{21}t_{2}} & \cdots & rs_{r_{21}t_{n}} \\
\vdots & \vdots & \ddots & \vdots \\
rs_{r_{mn}t_{1}} & rs_{r_{mn}t_{2}} & \cdots & rs_{r_{mn}t_{n}}
\end{bmatrix}$$

- Most direct visualization of changes in the domain
- Distinguish: changes in the domain versus changes in the user's perception.
Games with a Purpose

"We are headed down a path that is certain to end in the destruction of our experiment in democracy."
Evidence Confidence Matrix

- Records the per evidence source confidence between source and candidate terms and its temporal evolution → base for optimization and learning of the weights

\[
ECM_{C_s, C_c} = \begin{bmatrix}
  c_{es_1, t_1} & c_{es_1, t_2} & \cdots & c_{es_1, t_m} \\
  c_{es_2, t_1} & c_{es_2, t_2} & \cdots & c_{es_2, t_m} \\
  \vdots & \vdots & \ddots & \vdots \\
  c_{es_n, t_1} & c_{es_n, t_2} & \cdots & c_{es_n, t_m}
\end{bmatrix}
\]

- Supports trend detection and perception experiments
- Perception != Truth; example: food for kids (Kinder Milk Slice)
Source Impact Vector

-Contributes the per source impact weights for translating evidence sources into spreading activation weights.
  → Determines the impact of a single source.

\[ SIV_{t_i} = \begin{bmatrix} I_{es_1} & I_{es_2} & \cdots & I_{es_n} \end{bmatrix} \]

-Initial settings
  - Domain experts
  - heuristics and metrics such as Google Page Rank, etc.

-Gets adjusted based on user feedback
  → better alignment with the optimal relation strength matrix
Summary

- Ontologies and the corresponding instance information play a crucial role in improving Web Intelligence.
- Timely adaptation of ontologies to changes in the domain are still challenging.
- Presented an approach for aligning ontologies to user perception that allows:
  - Keeping ontologies up to date
  - Analyzing the **sources** of changes (evidence confidence matrix)
  - Detecting trends and patterns (evidence confidence matrix, source impact vector)
  - Making the reasons for change more traceable (evidence confidence matrix, source impact vector)
Interface Description
www.weblyzard.com/interface

Media Watch on Climate Change
www.ecoresearch.net/climate

Contact + Login Requests
albert.weichselbraun@htwchur.ch
info@weblyzard.com