Media Watch on Climate Change

Geospatial Web Technology for Accessing Environmental Online Resources

http://www.ecoresearch.net/climate
IDIOM Project

- **Scientific Partner:**
  - MODUL University Vienna
  - Technical University Graz
  - Vienna University for Economics and Business Administration (Coordinator)

- **Commercial Partner:**
  - Gентics
  - Prisma
  - Österreichwerbung

- **Project**
  - Total Cost: 773,113 EUR
  - WU-CM + WU-TLS: 316,200 EUR
Research Questions

• How widespread is content redundancy, and what influences content replication within and across social networks?

• Does media interactivity influence information diffusion? Can existing models such as hub-and-spoke, syndication and peer-to-peer explain this influence?

• Which content placement strategies increase the impact on the target audience and support self-reinforcing content propagation in virtual communities (viral marketing)?

• Impact of macroscopic information flows on environmental awareness and public opinion? What are appropriate methods to measure and model the extent, dynamics and latency of this process?
US Election 2004

Weekly Snapshots of International Web Coverage

**Post-Election Update** (Jan 5, 2005). Each week, the US Election 2004 Web Monitor compares attention and attitude towards the presidential candidates among online media, the Web sites of environmental organizations, and Fortune 1000 companies. Keywords grouped by political party and geographic region summarize the key issues associated with each candidate.

link: [Chart](#) | [News Media](#) | [Environmental Sector](#) | [Fortune 1000](#)

**Global Warming** (Oct 13, 2004). A devastating hurricane season and the Russian government’s recent decision to ratify the Kyoto Protocol have put climate change back in the limelight. But how important is the issue of global warming in the run for US presidency?

link: [Global Warming](#) | [Greenhouse Gas](#) | [Climate Change](#)

**First TV Debate** (Oct 2, 2004). An analysis of more than one million Web pages complements opinion polls and highlights the impact of the first TV debate between George W. Bush and John Kerry, with both parties profiting. John Kerry’s performance accelerated the Democrats’ prior gains in media attitude.

link: [Chart](#)

**Energy Policy** (Sep 28, 2004). What strategies have presidential candidates suggested to cope with record-level energy prices and increasing domestic demand? Public glossaries of the Energy Information Association, the US Department of Energy, and the California Energy Commission helped compile the lists of terms represented in this analysis.

link: [Renewable Energy](#) | [Fossil Fuels](#) | [Nuclear Energy](#)
US Election 2004 Web Monitor

Select poll:  
- ABC/Post  
- CBS/ NYT times  
- CNN/USA/Gallup

Latest data: 31 Oct

Bush: 48%
Kerry: 46%
Nader: 1%
Key event:

30 Sep
Bush and Kerry go head-to-head in the first of three televised debates.

Special Reports

- a bill by the way that calls for a mere 13 percent increase in funding for research into alternative energy -- as opposed to John Kerry's plan to double spending on alternative energy research (League of Conservation Voters). 😊

- the bill, which President Bush signed into law on October 22, 2004, includes only a one-year ptc extension and expands the eligible electricity resources to include geothermal energy, solar energy, open-loop biomass, small irrigation power, as well as municipal solid waste (Union of Concerned Scientists). 😊

- an ideological chasm separates the environmental policies of the two candidates, with President George W. Bush favoring more use of domestic coal and oil to cut dependence on middle east oil, while Kerry seeks a shift to clean energy like solar or wind power by 2020 (Environmental News Network). 😊

- they both talk about hydrogen as a fuel source and fuel cells, a technology for moving our cars and buses that President Bush has talked about and that John Kerry has made part of his campaign as well (Public Broadcasting Service). 😊

- "this is the first significant funding at the national level since President Bush's stated goal of supporting establishment of a hydrogen economy (ConocoPhillips). 😊

- the Texas RPS, passed in 1999 under Governor George W. Bush, has been so successful that utility companies have surpassed the percentages of renewable energy required by them (Friends Of The Earth). 😊
Web Mining Architecture
Media Watch on Climate Change

- **Public Web Portal**
  
  [http://www.ecoresearch.net/climate/](http://www.ecoresearch.net/climate/)

- **Extraction of Environmental Knowledge**
  
  - 150 Anglo-American News Media Sites
  - 200,000 Documents (Mirrored in Weekly Intervals)
  - 10,000 Environmental Articles

- **Contextualized Information Space**
  
  - **Spatial** Annotate Source and Target Geography
    
    BBC: „Vienna Marking Mozart Milestone“
    
    Source: Europe | United Kingdom | London
    
    Target: Europe | Austria | Vienna
  
  - **Semantic** Classify Documents, Assign Ontology Concepts
  
  - **Temporal** Add Timestamps to Publications, Events, etc.
Climate Change Media Watch

How to 'go green' on a budget - MSN Money

Published in: US • Text refers to: Portland • Mirror Date: 2007-07-04

Green logo green, logo more green content expand, Live Earth logo. Spinal Tap. Turn it up to 11. Spinal Tap is reuniting for Live Earth to prevent the earth from "spontaneously combusting." Live Earth concert information. Are democrats greener. * The green redneck. * Global warming survival handbook. * How to green your pet. Shopping for deals Corbis. Extra. How to 'go green' on a budget. Yes, organic food and products often cost more. Here are some ways to rethink your objectives and stretch your dollar. By Abby Schultz. The consumer wise grabs the half-gallon of organic milk instead of regular milk knows something about the cost of "going green." It's high. A recent study by the U.S. Department of Agriculture spells it out. Nationally, organic milk cost 95% more than regular milk in 2004: $4.01 for organic, $2.02 for regular, a price difference of $1.99. That's almost enough to buy another half-gallon of regular milk. Prices of organic produce, poultry and eggs are also higher, far more than 200% higher in the case of poultry, according to the USDA. But green-minded consumers -- who care about their health, as well as "green"
Interface Services

- Semantic Interfaces
  - Perceptual Maps (US Election Monitor 2004)
  - Information Landscapes (Aureka, VisIslands)
  - Ontology-based Visualizations
Perceptual Maps
Thematic Maps

Thematic Maps

InfoSky

Aureka

VisIslands
Ontology Services
Interface Services

- Semantic Interfaces
  - Perceptual Maps (US Election Monitor 2004)
  - Information Landscapes (Aureka, VisIslands)
  - Ontology-based Visualizations

- Geospatial Interfaces
  The Geospatial Web “may ultimately be the big disruptive innovation of the coming decade” (Erle et al. 2005, xxv).
  
  - 2D Platforms
    - MapQuest
    - Google Maps
    - Yahoo! Maps
  - 3D Platforms
    - NASA World Wind
    - Google Earth
    - MS Virtual Earth
## Most frequently searched keywords

**Region/websites:** All  
**Languages:** All  
**Period:** 1/1/2006-30/4/2006

<table>
<thead>
<tr>
<th>The most frequent 20 words</th>
<th>n</th>
<th>in %</th>
<th>Results</th>
<th>in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 wien</td>
<td>1.562</td>
<td>1,3</td>
<td>315,544</td>
<td>8,4</td>
</tr>
<tr>
<td>2 salzburg</td>
<td>1.422</td>
<td>1,2</td>
<td>175,327</td>
<td>4,6</td>
</tr>
<tr>
<td>3 vienna</td>
<td>1.340</td>
<td>1,1</td>
<td>262,324</td>
<td>7,0</td>
</tr>
<tr>
<td>4 innsbruck</td>
<td>1.282</td>
<td>1,1</td>
<td>102,051</td>
<td>2,7</td>
</tr>
<tr>
<td>5 graz</td>
<td>1.113</td>
<td>0,9</td>
<td>60,270</td>
<td>1,6</td>
</tr>
<tr>
<td>6 linz</td>
<td>794</td>
<td>0,7</td>
<td>22,637</td>
<td>0,6</td>
</tr>
<tr>
<td>7 hotel</td>
<td>775</td>
<td>0,6</td>
<td>53,101</td>
<td>1,4</td>
</tr>
<tr>
<td>8 mozart</td>
<td>756</td>
<td>0,6</td>
<td>62,654</td>
<td>1,7</td>
</tr>
<tr>
<td>9 ischgl</td>
<td>742</td>
<td>0,6</td>
<td>10,945</td>
<td>0,3</td>
</tr>
<tr>
<td>10 kufstein</td>
<td>692</td>
<td>0,6</td>
<td>9,351</td>
<td>0,2</td>
</tr>
<tr>
<td>11 zell am see</td>
<td>653</td>
<td>0,5</td>
<td>40,649</td>
<td>1,1</td>
</tr>
<tr>
<td>12 hotels</td>
<td>592</td>
<td>0,5</td>
<td>28,575</td>
<td>0,8</td>
</tr>
<tr>
<td>13 zürs</td>
<td>575</td>
<td>0,5</td>
<td>15,437</td>
<td>0,4</td>
</tr>
<tr>
<td>14 saalbach</td>
<td>493</td>
<td>0,4</td>
<td>7,591</td>
<td>0,2</td>
</tr>
<tr>
<td>15 tirol</td>
<td>492</td>
<td>0,4</td>
<td>74,665</td>
<td>2,0</td>
</tr>
<tr>
<td>16 seefeld</td>
<td>463</td>
<td>0,4</td>
<td>9,622</td>
<td>0,3</td>
</tr>
<tr>
<td>17 heiligenblut</td>
<td>457</td>
<td>0,4</td>
<td>3,437</td>
<td>0,1</td>
</tr>
<tr>
<td>18 camping</td>
<td>437</td>
<td>0,4</td>
<td>4,216</td>
<td>0,1</td>
</tr>
<tr>
<td>19 kaprun</td>
<td>429</td>
<td>0,4</td>
<td>5,148</td>
<td>0,1</td>
</tr>
<tr>
<td>20 wellness</td>
<td>420</td>
<td>0,4</td>
<td>20,321</td>
<td>0,5</td>
</tr>
</tbody>
</table>

**Total:** 119,588 100,0 3,772,664 100,0

Table: TW-X3
The Geospatial Web

- Integrates
  - Cartographic Data
  - Environmental Indicators
  - Geotagged Hypermedia

- Knowledge Planets
  - Topology of Information Landscape
    - Peak = Cluster of Documents on a Specific Topic
    - Valley = Sparsely Populated Part of the Information Space
  - Projecting Information Landscapes onto Virtual Globes
    - Initial Arrangement and Stability of Topics
    - Navigation Across Layers and the 0° Meridian Line

- Download | www.geospatialweb.com/chapter-1

Analytical Services
Processing Techniques

• Infrastructure
  • Virtualization (XEN)
  • Host Systems
    Siemens RX200/300 Server mit je 8 CPU's und 32 GB RAM
  • Eases administrative tasks
  • PC7 cluster

• Tasks
  • Mirroring
  • Tagging (Spatial, Temporal, Named Entities)
  • Semantic Maps
  • Ontology Extension
Distributed Computing – hadoop

- Developed by the Apache project (Lucene subproject)
- Implements MapReduce + a distributed Filesystem
Distributed Computing – hadoop

xmbalrog Hadoop Map/Reduce Administration

State: RUNNING
Version: 0.15.0. r589881
Compiled: Mon Oct 28 15:01:10 PDT 2007 by cutting
Identifier: 200711071257

Cluster Summary

<table>
<thead>
<tr>
<th>Maps</th>
<th>Reduces</th>
<th>Tasks/Node</th>
<th>Total Submissions</th>
<th>Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>24</td>
</tr>
</tbody>
</table>

Running Jobs

Running Jobs
none

Completed Jobs

<table>
<thead>
<tr>
<th>Jobid</th>
<th>User</th>
<th>Name</th>
<th>Map % Complete</th>
<th>Map Total</th>
<th>Maps Completed</th>
<th>Reduce % Complete</th>
<th>Reduce Total</th>
<th>Reduces Completed</th>
</tr>
</thead>
<tbody>
<tr>
<td>job_200711071257_0001</td>
<td>aweichae</td>
<td>streamjob23908.jar</td>
<td>100.00%</td>
<td>26</td>
<td>26</td>
<td>100.00%</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Failed Jobs

Failed Jobs
Keep things simple

- Overhead ↓ -> Processing Time ↓
- Avoid costly operations
  - Database writes, Networking, IO
- DFS Input Data -> DFS Output Data
- NFS about twice as fast
  - file names
  - Handle output of redeployed tasks
Use your Database and Data Structures wisely

• **INSERT vs. COPY**
  • Example: 4 million DB writes (prepared statements, stored procedures, ...)
    - 2.5 days without results vs.
    - 30 seconds(!) + 2 min (COPY)
  • Constraints and Indices

• **Cache data (!)**
  • Example: Computation of the Semantic Map
    - 3 hours -> 18 min (10x)

• **Use prepared statements (DB Planer)**
  • Approx. 10x

• **Data Structures:**
  e.g. ArrayLists vs. HashSet; List vs. Set