KPIs from Web Agents for Policies’ Impact Analysis and Products’ Brand Assessment

Antonio Candiello – Agostino Cortesi

Computer Information Systems & Industrial Management Applications

Agent-Based Computing

DAIS – Dipartimento di Scienze Ambientali, Informatica e Statistica, Università “Ca’ Foscari” di Venezia

work partially supported by Direzione Sistemi Informativi

CISIM 2011 – 10th International Conference on Agent-Based Computing
Summary

1. **A PDCA Model for eGovernment**
   - the approach
   - the *eGovernment Intelligence* framework
   - the software modules

2. Technology & Maps
   - *webbots*, data scrapers, spiders
   - the open source *SpagoBI* engine
   - accumulating data in DBs

3. **Ongoing Research**: ICT-related gender gap KPIs
Main Goals

Explicit a comprehensive model for:

- the management of Enterprises’ marketing initiatives or PAs’ innovation projects,
- their systematic monitoring and related impact analysis measurement
- supported by an integrated Web Agents Intelligence Information System capable of:
  - registering and monitoring such policies,
  - monitor the relative initiatives/projects against their goals,
  - systematically evaluating their impact,
  - reviewing the policies themselves.
PAs and Enterprises

The information system developed within this applied research provides a continuous source for:

- **Public administrators**, that have the capability to
  - *continuously improve* their services via an *objective evaluation of the resulting impact*,
  - keeping their *citizens better informed* and up-to-date regarding *goals* set in advance for the policies and the *success rate* of the local government funded innovation initiatives carried out for the public benefit.

- **Enterprise managers**, on the other hand,
  - are able to *evaluate* the *effective appreciation* of their products/brands by the consumers;
  - can *inspect* *specific attributes* that customers associate to the products (like *performance*, *durability*, *comfort*).
Sub-objectives

a) finding an **objective validation** for the effectiveness of PAs / Enterprises policies,

b) **qualifying** and **quantifying** the **effectiveness** through appropriate impact statistical territorial **indicators**,  
c) gathering the relevant indicators via **automatic webbots/scrapers** and **semiautomatic extractors/wrappers**, completing the data when needed with focused survey campaigns,  
d) **representing** and **mapping** the indicators showing the **explicit relation with the affecting innovation projects** and the areas involved.
Classes of indicators

- **direct innovation** indicators, mainly the ICT indicators *enabling the Information Society*
- **indirect socio-economical** indicators related to the *resultant impact* of the innovation over the society
- **specific product-** and **brand-related** indicators able to report the evidence of consumers *positive or negative opinions* regarding specific products or lines of products.
The Approach

1. Adapt the **PDCA improvement cycle** for Enterprises’ marketing initiatives / PA innovation projects

2. Harvest the “**Raw Web**” with *webbots, data scrapers, crawlers, spiders*, less reliable but frequently updated

3. Use validated **official data** from Public Authorities and Official Institutions via the “**Deep Web**” (online DBs)

4. *In-progress research*: use the “**Linked Data**” of *semantic web*, access the “Open Government Data”

5. **Geo-refentiate** the data to the *municipalities* level

6. **Store** the KPI obtained **every day** (sort of “KPI Wayback Machine”)

---

Calcutta, India, 14-16/12/2011

Dipartimento di Scienze Ambientali, Informatica e Statistica, Università “Ca’ Foscari” di Venezia

KPIs from Web Agents …

A. Candiello & A. Cortesi
A PDCA Model for Enterprises’ mkt initiatives & PA innovation projects

**Policy Manager**

*Plan:* define local policies, select innovation KPIs, set goals.

*Do:* run projects, begin measurements, monitor KPIs.

**Web Agents Intelligence**

*Act:* review policies, redefine goals.

*Check:* analyse indicators, verify goal achievements, view georeferentiate data.

**Policy Manager**

**Event Scheduler**

**Agents Manager**

**Map Viewer**
DO

CHECK

ACT

PLAN

Policy Manager

KPIs

Targets

Events Scheduler

Agents Manager

Map Viewer

Policy Manager

Plan: define local policies, select innovation KPIs, set goals.

Act: review policies, redefine goals.

Check: verify goal achievements, view geo-referentiated data.

Do: run projects, begin measurements, monitor KPIs.

to define policies, projects, actions and to set goals

to modify the policies & projects

to visualize the data via maps & tables

to timely wake-up the daemons

to manage webbots, adapters & data scrapers

KPIs' db

Targets

HF channel
high frequency

LF channel
low frequency

ok _____

ko _____
Web Intelligence for Enterprises mkt initiatives & PAs’ innovation projects

Definition
Policies for Marketing or Innovation

Intervention
Start of Marketing Initiatives / Innovation Projects

Verification
Innovation Impact Analysis & Product Brand Assessment

Assessment
Review of Policies

Policy Management

Web Agents Intelligence

Targeting
Choice of KPIs & Goals / Thresholds

Monitoring
KPIs’ staggered measurement

Analysis
Geo-Mapping & Statistical Analysis

Evaluation
Identification of Critical Points

Calcutta, India,
14-16/12/2011
Dipartimento di Scienze Ambientali, Informatica e Statistica, Università "Ca' Foscari" di Venezia
KPIs from Web Agents .10
A. Candiello & A. Cortesi
**Intervention**

Start of Marketing Initiatives / Innovation Projects

**Monitoring**

KPIs' staggered measurement

- Online DBs
- Data adapters
- Sources
- Direct feed
- LF channel
- Low frequency

- Spiders & webbots
- Raw web
- Indirect feed
- HF channel
- High frequency
- Keywords

- Citizens feedback on ICT use & services
- Surveys on ICT
- Surveys for campaigns

- KPI Database
- Staggered data consolidation

**Plan**

**Do**

**Act**

**Check**

**KPIs from Web Agents**

A. Candiello & A. Cortesi

Calcutta, India, 14-16/12/2011

Dipartimento di Scienze Ambientali, Informatica e Statistica, Università "Ca' Foscari" di Venezia
DIPARTIMENTO DI SCIENZE AMBIENTALI, INFORMATICA e STATISTICA

Verification
Innovation Impact Analysis & Product Brand Assessment

Analysis
Geo-Mapping & Statistical Analysis

Web Extractors
Low Frequency (LF), High Reliability Channel
Semi-automatic

Webbots & Scrapers
High Frequency (HF), Low Reliability Channel
Automatic

Surveys
Highly Focused but Costly, Medium Reliability Channel
Manual / partly web assisted

geo-visualization
layer model
business intelligence
KPI dbms

Calcutta, India, 14-16/12/2011
Dipartimento di Scienze Ambientali, Informatica e Statistica, Università "Ca' Foscari" di Venezia
A. Candiello & A. Cortesi

KPIs from Web Agents: 38
Software Modules

- **Policy Manager** [PLAN],
  
to input/define policies, projects, indicators, targets,

- **Events Scheduler** [DO],
  
for the reliable planning of monitoring events with (day as minimal temporal unit),

- **Agents Manager** [DO],
  
managing webbots, adapters & scrapers,

- **Map Viewer** [CHECK],
  
to visualize the data via maps & tables via SpagoBI,
**KPIs from data sources and their representation**

- **Currently:**
  - “Official” Data collected via adapters/extractors or online webbots. **High quality data.**
  - “Raw” Data collected via webbots/scrapers/spiders on the web (the easiest: # of results of searches). **Low quality data.**

- **In the next future:**
  - Open Government Data (eGovernment),
  - Open Linked Data (semantic web)
  - also: Local Government could also collect data from different (& raw) web data sources, validate data and expose. Needed data quality assurance activity

- Data is accumulated **daily/weekly/monthly** in a DB (**twitter scraping** would need higher frequencies)
- Geo-referentiation of data: at regional, provincial, **municipality** level
Advantages of technology used

- **Java** (also used **PHP** and **Python**) for **data daemons** (webbots/scrapers/adapters)
  - specific **webbots/scrapers** for each social or business network/community; needed elaborations of data retrieved. Data updated **daily**. Host sites **could discourage** access to web robots **difficult to maintain**
  - specific **adapters** for each Institutional online or offline (e.g. **csv** files) data source. Eurostat offers a **wide** set of formats & access modalities. However: data updated **yearly or monthly** **easy to maintain**

- **SpagoBI** to represent and interact with the data on the maps
  - **Public Administrators** and **Enterprise Managers** appreciates transposition of **data on maps**
  - **Patterns of effectiveness** of **policies** are **clearly visible**

- **Postgres DB** to deposit the data
  - Accumulating data makes possible to analyse the growth **trends**
Accessing the big Web Repositories

- **Youtube**
  - used the **Youtube APIs** for extracting KPIs on video production & consume (#uploads, #views)
  - also: web parsing youtube site

- **Facebook**:
  - **not using Facebook APIs** (as users have to agree to use the application): we extract data with general web parsing techniques

- **Google**:
  - **not using Google Search APIs** (limited to max 100 daily searches)
  - using **Google Analytics APIs** for (owned/managed) web site accesses
  - **Limits** on number of searches that could be done in a period of time.

- **Yahoo**:
  - **not using Yahoo! Search APIs** (limited to max 100 daily searches)
  - using **Yahoo Sites APIs** for web sites size and relevance (#pages, #inlinks)
  - **Limits** on number of searches that could be done in a period of time.

- **General Web**:
  - **HTML Cleaner APIs** for (bad-formed HTML) web parsing (for Java)
  - used **Schrenk’s webbots library** for PHP
  - **Blogs** and **forums** are the more interesting sources of information (frequently updated and massive)
Mapping the data on the territory: 
content production & consume

Youtube – **uploads**
*(producers, inhomogeneous)*

Youtube – **views**
*(consumers, more homogeneous)*

Calcutta, India, 14-16/12/2011

Dipartimento di Scienze Ambientali, Informatica e Statistica, Università "Ca' Foscari" di Venezia

KPIs from Web Agents 38
A. Candiello & A. Cortesi
Mapping the data on the territory: Work & Education

# of ICT Enterprises (work)

# of Schools (education)
# Mapping the data on the territory: ICT relevance (# of hits)
An example for brand assessment

Nielsen *Brand Association Map* (BAM)

<table>
<thead>
<tr>
<th></th>
<th>Cool</th>
<th>Confort</th>
<th>Expensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adidas</td>
<td>103</td>
<td>45</td>
<td>12</td>
</tr>
<tr>
<td>Nike</td>
<td>198</td>
<td>106</td>
<td>338</td>
</tr>
<tr>
<td>Rebook</td>
<td>31</td>
<td>19</td>
<td>5</td>
</tr>
<tr>
<td>Converse</td>
<td>54</td>
<td>28</td>
<td>16</td>
</tr>
</tbody>
</table>

Googling with the words, e.g. 
“+Adidas +cool”
### Example KPIs (Veneto Region, municipalities > 25.000 inhabitants)

<table>
<thead>
<tr>
<th>comune</th>
<th>Population</th>
<th>Area (km²)</th>
<th>Total Income</th>
<th>Incomes</th>
<th># of Schools</th>
<th>n.video.youtube</th>
<th>n.views.youtube</th>
<th>n.hits.youtube</th>
<th>n.pages.egov</th>
<th>n.inlinks.egov</th>
<th>n.ICTenterprises</th>
<th>% wide band</th>
</tr>
</thead>
<tbody>
<tr>
<td>Venezia</td>
<td>270.801</td>
<td>413</td>
<td>€ 4.089.444.597</td>
<td>183.323</td>
<td>281</td>
<td>1.417</td>
<td>450.274</td>
<td>7.750.000</td>
<td>232.772</td>
<td>60.572</td>
<td>146</td>
<td>100,0%</td>
</tr>
<tr>
<td>Verona</td>
<td>264.475</td>
<td>207</td>
<td>€ 4.030.822.293</td>
<td>153.339</td>
<td>278</td>
<td>1.453</td>
<td>396.203</td>
<td>2.520.000</td>
<td>14.033</td>
<td>20.525</td>
<td>237</td>
<td>100,0%</td>
</tr>
<tr>
<td>Padova</td>
<td>212.989</td>
<td>93</td>
<td>€ 3.596.976.438</td>
<td>123.422</td>
<td>216</td>
<td>1.355</td>
<td>177.417</td>
<td>3.500.000</td>
<td>2.880</td>
<td>6.620</td>
<td>311</td>
<td>100,0%</td>
</tr>
<tr>
<td>Vicenza</td>
<td>115.550</td>
<td>81</td>
<td>€ 1.719.538.821</td>
<td>65.709</td>
<td>128</td>
<td>1.269</td>
<td>479.754</td>
<td>2.620.000</td>
<td>36.230</td>
<td>43.927</td>
<td>132</td>
<td>100,0%</td>
</tr>
<tr>
<td>Treviso</td>
<td>82.208</td>
<td>56</td>
<td>€ 1.405.295.466</td>
<td>48.569</td>
<td>103</td>
<td>1.248</td>
<td>131.630</td>
<td>2.570.000</td>
<td>8.397</td>
<td>9.883</td>
<td>110</td>
<td>100,0%</td>
</tr>
<tr>
<td>Rovigo</td>
<td>52.118</td>
<td>109</td>
<td>€ 747.269.908</td>
<td>31.361</td>
<td>64</td>
<td>1.052</td>
<td>35.704</td>
<td>1.770.000</td>
<td>8.180</td>
<td>5.872</td>
<td>60</td>
<td>95,0%</td>
</tr>
<tr>
<td>Chioggia</td>
<td>50.772</td>
<td>185</td>
<td>€ 513.947.858</td>
<td>25.127</td>
<td>44</td>
<td>24</td>
<td>6.707</td>
<td>78.900</td>
<td>4.804</td>
<td>2.081</td>
<td>28</td>
<td>92,0%</td>
</tr>
<tr>
<td>Bassano del Grappa</td>
<td>43.015</td>
<td>47</td>
<td>€ 590.078.083</td>
<td>23.737</td>
<td>50</td>
<td>151</td>
<td>20.301</td>
<td>154.000</td>
<td>5.630</td>
<td>4.216</td>
<td>52</td>
<td>100,0%</td>
</tr>
<tr>
<td>San Dona di Piave</td>
<td>41.247</td>
<td>79</td>
<td>€ 541.208.071</td>
<td>23.339</td>
<td>31</td>
<td>87</td>
<td>4.772</td>
<td>103.000</td>
<td>6.659</td>
<td>1.703</td>
<td>23</td>
<td>95,0%</td>
</tr>
<tr>
<td>Schio</td>
<td>39.586</td>
<td>67</td>
<td>€ 538.858.210</td>
<td>23.305</td>
<td>35</td>
<td>197</td>
<td>17.680</td>
<td>881.000</td>
<td>9.597</td>
<td>7.216</td>
<td>34</td>
<td>100,0%</td>
</tr>
<tr>
<td>Mira</td>
<td>38.857</td>
<td>99</td>
<td>€ 499.994.000</td>
<td>22.727</td>
<td>28</td>
<td>1.375</td>
<td>2.310.000</td>
<td>162</td>
<td>3.071</td>
<td>36</td>
<td>90,0%</td>
<td></td>
</tr>
<tr>
<td>Belluno</td>
<td>36.618</td>
<td>147</td>
<td>€ 577.581.894</td>
<td>23.157</td>
<td>49</td>
<td>1.104</td>
<td>66.209</td>
<td>1.310.000</td>
<td>2</td>
<td>5.374</td>
<td>38</td>
<td>100,0%</td>
</tr>
<tr>
<td>Conegliano</td>
<td>35.676</td>
<td>36</td>
<td>€ 530.114.920</td>
<td>20.611</td>
<td>55</td>
<td>245</td>
<td>60.795</td>
<td>181.000</td>
<td>11.958</td>
<td>6.484</td>
<td>35</td>
<td>100,0%</td>
</tr>
<tr>
<td>Castelfranco Veneto</td>
<td>33.675</td>
<td>51</td>
<td>€ 454.959.141</td>
<td>18.907</td>
<td>53</td>
<td>93</td>
<td>41.112</td>
<td>146.000</td>
<td>1</td>
<td>146</td>
<td>36</td>
<td>100,0%</td>
</tr>
<tr>
<td>Villafranca di Verona</td>
<td>32.866</td>
<td>24</td>
<td>€ 419.241.447</td>
<td>18.657</td>
<td>11</td>
<td>545</td>
<td>60.065</td>
<td>92.200</td>
<td>534</td>
<td>70</td>
<td>43</td>
<td>100,0%</td>
</tr>
<tr>
<td>Montebelluna</td>
<td>30.948</td>
<td>49</td>
<td>€ 406.756.610</td>
<td>17.342</td>
<td>34</td>
<td>110</td>
<td>10.236</td>
<td>89.000</td>
<td>4.753</td>
<td>4.090</td>
<td>41</td>
<td>100,0%</td>
</tr>
<tr>
<td>Vittorio Veneto</td>
<td>29.210</td>
<td>83</td>
<td>€ 395.514.950</td>
<td>17.159</td>
<td>45</td>
<td>207</td>
<td>9.651</td>
<td>611.000</td>
<td>4.173</td>
<td>1.542</td>
<td>45</td>
<td>91,0%</td>
</tr>
<tr>
<td>Mogliano Veneto</td>
<td>28.125</td>
<td>46</td>
<td>€ 430.201.499</td>
<td>16.393</td>
<td>28</td>
<td>149</td>
<td>41.962</td>
<td>27.300</td>
<td>3.679</td>
<td>2.265</td>
<td>31</td>
<td>95,0%</td>
</tr>
<tr>
<td>Valdagno</td>
<td>26.829</td>
<td>50</td>
<td>€ 362.690.849</td>
<td>17.208</td>
<td>28</td>
<td>126</td>
<td>68.010</td>
<td>19.300</td>
<td>12.107</td>
<td>2.897</td>
<td>18</td>
<td>90,0%</td>
</tr>
<tr>
<td>Mirano</td>
<td>26.795</td>
<td>46</td>
<td>€ 383.648.244</td>
<td>15.722</td>
<td>30</td>
<td>286</td>
<td>12.547</td>
<td>203.000</td>
<td>3.204</td>
<td>699</td>
<td>137</td>
<td>100,0%</td>
</tr>
<tr>
<td>Spinea</td>
<td>26.674</td>
<td>15</td>
<td>€ 371.654.778</td>
<td>16.056</td>
<td>20</td>
<td>5</td>
<td>1.337</td>
<td>40.600</td>
<td>5.500</td>
<td>976</td>
<td>40</td>
<td>100,0%</td>
</tr>
<tr>
<td>Arzignano</td>
<td>25.823</td>
<td>34</td>
<td>€ 328.905.490</td>
<td>13.837</td>
<td>22</td>
<td>155</td>
<td>14.227</td>
<td>87.300</td>
<td>2.138</td>
<td>754</td>
<td>62</td>
<td>100,0%</td>
</tr>
<tr>
<td>Legnago</td>
<td>25.556</td>
<td>80</td>
<td>€ 339.314.492</td>
<td>14.956</td>
<td>33</td>
<td>223</td>
<td>12.599</td>
<td>76.300</td>
<td>3.208</td>
<td>750</td>
<td>34</td>
<td>100,0%</td>
</tr>
<tr>
<td>Portogruaro</td>
<td>25.406</td>
<td>102</td>
<td>€ 351.142.820</td>
<td>15.635</td>
<td>30</td>
<td>166</td>
<td>23.712</td>
<td>181.000</td>
<td>3.657</td>
<td>18.889</td>
<td>26</td>
<td>90,0%</td>
</tr>
<tr>
<td>Jesolo</td>
<td>25.232</td>
<td>95</td>
<td>€ 301.490.304</td>
<td>14.684</td>
<td>20</td>
<td>681</td>
<td>63.013</td>
<td>140.000</td>
<td>2.919</td>
<td>103</td>
<td>20</td>
<td>100,0%</td>
</tr>
</tbody>
</table>

**Signs of good community vitality**

- KPIs from Web Agents
- Work partially supported by A. Candiello & A. Cortesi

**Bad data (name match)**

- J.Karels
- Bad data (name match)
- Delta KPIs from Web Agents

---

**Calcutta, India, 14-16/12/2011**

Dipartimento di Scienze Ambientali, Informatica e Statistica, Università "Ca' Foscari" di Venezia
### Youtube KPIs: video *uploads & views* (Belluno province, municipalities)

<table>
<thead>
<tr>
<th>Municipality</th>
<th>Uploads</th>
<th>Views</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belluno</td>
<td>506</td>
<td>35035</td>
</tr>
<tr>
<td>Feltre</td>
<td>75</td>
<td>36556</td>
</tr>
<tr>
<td>Santa Giustina</td>
<td>56</td>
<td>30088</td>
</tr>
<tr>
<td>Borca di Cadore</td>
<td>53</td>
<td>6533</td>
</tr>
<tr>
<td>Sedico</td>
<td>52</td>
<td>10061</td>
</tr>
<tr>
<td>Agordo</td>
<td>37</td>
<td>17231</td>
</tr>
<tr>
<td>Cesiomaggiore</td>
<td>26</td>
<td>29046</td>
</tr>
<tr>
<td>Domegge di Cadore</td>
<td>21</td>
<td>6976</td>
</tr>
<tr>
<td>Pedavena</td>
<td>17</td>
<td>26246</td>
</tr>
<tr>
<td>Mel</td>
<td>17</td>
<td>26748</td>
</tr>
<tr>
<td>Longarone</td>
<td>12</td>
<td>74633</td>
</tr>
<tr>
<td>Sappada</td>
<td>11</td>
<td>11225</td>
</tr>
<tr>
<td>Limana</td>
<td>11</td>
<td>6953</td>
</tr>
<tr>
<td>Trichiana</td>
<td>10</td>
<td>8695</td>
</tr>
<tr>
<td>Selva di Cadore</td>
<td>10</td>
<td>8960</td>
</tr>
<tr>
<td>Lozzo di Cadore</td>
<td>10</td>
<td>2015</td>
</tr>
</tbody>
</table>
Different KPI layers (Belluno province)

Population

Search Engine, ICT-related

Income

Youtube uploads

KPIs from Web Agents
Thanks for your attention
Antonio Candiello, Agostino Cortesi

DAIS – Dipartimento di Scienze Ambientali, Informatica e Statistica,
Università “Ca’ Foscari” di Venezia
email candiello@dsi.unive.it