The issue of IQ in internet-based early-warning systems for trend management

Daniel Diemers
University of St. Gallen HSG, Switzerland
(Research Assistant)
Wedoso gmbh
(Managing Partner)
SfS-HSG, Ackerstrasse 44
8005 Zürich, Switzerland
e-mail: daniel@diemers.net

Practice-oriented Paper

Executive Summary

This contribution presents our latest experiences with internet-based early-warning systems that are used in firms in different areas with different functionalities to facilitate trend management and risk management. In these systems, intelligent software agents are deployed to gather and analyze information and monitor sites on the internet.

A central problem in this context is to calibrate these systems in a way that data mining and analysis processes deal with all mission-critical data available, but are still able to operate with a manageable amount of data. Thus, IQ plays a central role in measuring the performance of such a system and in achieving a sound balance between quality information and information overload.

Our methodology includes a distinct framework of IQ and respective procedures and algorithms that are based on a socially biased, community-oriented perspective of IQ that has been presented in past contributions to this conference.
“The issue of IQ in internet-based early-warning systems for trend management”

Dr. Daniel Diemers
University of St. Gallen (HSG), Switzerland
Institute of Sociology SfS

Structure
1. Research Context: Trend Management
2. Internet-based Early Warning Systems
3. Our Framework for IQ
4. Practical Experiences with IQ issues
5. Conclusions and open questions

1. Research Context

• Contingency Management, Knowledge Management, Early-Warning Systems
• Research Question:
  – „How can we measure and quantify the „quality“ of an identified Web source within an early-warning system?”

10/5/01 dr. daniel diemers 3

Contingency Management (CM)

Locating EWS in a CM Framework

• Internal view:
  – Knowledge Management, Data Mining, MIS, Business Intelligence
• External View:
  – Trend Scouts, Issue Management, Clipping Services, Monitoring Service, Search Engines
• soft-skill vs. IT-based Methodology
• future/external vs. past/internal data

10/5/01 dr. daniel diemers 5

Proceedings of the Sixth International Conference on Information Quality
2. Early Warning Systems

- Early Warning Systems shall help firms in their perception of the contingent corporate environment
- Early Warning Systems will play an increasingly important role in coping with the volatility and dynamics of markets

a feasible EWS Methodology

- 3 Phases:
  - Focused Crawling: Community Topography
  - Scanning/Monitoring/Analyzing
  - Border Control and Topography Updating
- 3 Modules:
  - software-agents
  - analysis module
  - reporting & iteration

The Community Perspective

- Diemers Theorem: "only information that is referenced to or embedded in the respective community can develop any potential relevance".
- Identify Hubs & Authorities within topical communities

Why Community Knowledge?

- all available information in virtual communities on a given subject
- monitored community on a given subject

Modelling Virtual Spaces

Static Methodology Overview
3. Our Framework for IQ

- How can we measure and quantify the "quality" of an identified Web source within an early-warning system?
- Agent-based IQ assessment process
- 3 levels of IQ criterias applied

3 Levels of IQ Criterias

- General IQ Criteria (analysis)
- First Order Relevancy (semantic analysis)
- Second Order Relevancy (feedback loop)

IQ Quantification Model

IQ Cat. I: General IQ Criteria

1. Latency (Agents sleep/frozen) #seconds
2. Server Performance #Kbits/s
3. Relative Size of Source #Kbyte/Average
4. Age of Site #days
5. Age of Last Update #days
6. IP Address / Domain Name semantic rating
IQ Cat. II: First Order Relevancy

1 Keyword matches
   #matches
   analysis according to semantic networks with example text and keywords. Certain value triggers second order analysis
2 Number of Links
   #links
   number of links leaving site indicates hub quality

Cat. III: Second Order Relevancy

1 Mentioning/Reference of Site
   #
   AUTHORITY
2 Number of inside links incoming
   #links
   AUTHORITY
3 Number of inside links outgoing
   #links
   HUB
4 Site already analyzed
   #days
5 Similarity of Site
   %
6 Reassessment of Relevancy
   semantic analysis

IQ Quantification Model

Site „xy.net“:

REL [0.65] HUB [0.24] AUTH[0.67]

Cat. I Cat. II Human

Example:

IQ Criteria I.6: IP/Domain Address

• Semantic analysis and rating
  – type of primary address and host:
    www.relevant.org/index.htm
    www.member.tripod.net~john/xy.htm
    www.freenet.net/classof98page/xy.htm
  – domain name extensions
    .edu, .gov, .com, .biz, .net, .to, .li, ...etc.

Example:

Realtime IQ Measuring

4. Practical Experiences

• Achieving the right balance between Information Overload and Capturing all relevant Information
• Getting human IQ input into the early-warning system vs. automating the routine agent processes
• Applying high-quality semantic analysis (which theory/technique? which tools?)
Practical Experiences (cont.)

• Establish a learning-system, especially in respect to IQ
• In Scanning Mode: achieve „minimum tolerance“, 90% „no blind-spots“ reliability
• In Monitoring Mode: achieve „zero tolerance“, 100% speed and reliability

5. Conclusions

• Internet-based early-warning systems are increasingly becoming accepted and useful tools for management
• monitoring and scanning virtual spaces requires highly sophisticated methodology and agent/software technology
• reliable results depend mainly on a sound and applicable framework for IQ

IQ’2001 community!
Thank You For Your Attention!

contact information:
dr. daniel diemers
Institute of Sociology SFS, University of St. Gallen (HSG), Switzerland
Partner, Wedoso GmbH, Zurich, Switzerland
Ackerstrasse 44, 8005 Zurich, Switzerland
+41 1 271 18 35, e-mail: daniel@diemers.net