Online Communities

With slides by ROBUST Consortium and in particular Harith Alani

Steffen Staab
High Level Objectives

- Risk Management
  - Risk modelling
  - Detection
  - Automatic reaction

- Community Analysis
  - Contents
  - Single users
  - Entire communities

- Community Forecasting
  - Policies
  - Prediction
  - Decision support

- Large Scale Processing
  - Big Data
  - Realtime
  - Parallel Processing
WP7-9: Use Cases

Lotus Connections
Communities
• Employees
• Working groups
• Interest Groups
• Projects

Business value
• Task relevant information
• Collaboration
• Innovation

Volume
• 4,000 posts/day
• 386,000 employees
• 1.5GB content/day

IBM

SAP Community Network (SCN)
Communities
• Customers
• Partners
• Suppliers
• Developers

Business value
• Products support
• Services
• Find business partners

Volume
• 6,000 posts/day
• 1,700,000 subscribers
• 16GB log/day

SAP

MeaningMine
Communities
• Social media
• News
• Web fora
• Public communities

Business value
• Topics
• Opinions
• Service for partners

Volume
• 1,400,000 posts/day
• 708,000 web sources
• 45GB content/day

Polecat

Employees
• Intranet

Business Partners
• Extranet

Public Domain
• Internet

Volume
• 6,000 posts/day
• 1,700,000 subscribers
• 16GB log/day

Community

Volume
• 4,000 posts/day
• 386,000 employees
• 1.5GB content/day

Community

Volume
• 6,000 posts/day
• 1,700,000 subscribers
• 16GB log/day

Community

Volume
• 1,400,000 posts/day
• 708,000 web sources
• 45GB content/day
SIOC – Semantically interlinked Online Communities

Blogosphere
- Weblog
- Post
- Weblog
- Post

Boardscape
- Bulletin Board
- Post
- Post
- Post

Usenetverse
- Usenet Group
- Post

Listsphere
- Mailing List
- Post
- Post
- Post

Virtual Forum
- Distributed Conversation

Unified Community

7 years of experience representing communities

http://west.uni-koblenz.de
Overview

Analysis and Management of Individuals and Communities

Risks and Opportunities
- Policies and rules for intervention
- Risk templates
- Risk Management Dashboard
- Risk Tracking

Individuals and Behaviours
- Parameters Influencing User Behaviour
- Model of User Needs
- User Classification

Communities and Topics
- Ontology of Behaviour Types
- Cross-community Dynamics
- Topic and Discussion Tracking
- Community Structure Analysis

Event Bus, Query and Information Exchange

WP1
WP2
WP3
WP4
WP5

Extract Transform Load Tools
Parallel Data Processing Platform
Agent-Based Community Simulation
Mathematical Models of Communities
Algebra and Operators for Analytics
Simulation Platform

Online Business Communities
Social Network Platform
Content
Log files

High Performance Data Processing and Community Simulation

Introduction to Web Science
http://west.uni-koblenz.de
Slide 6 of 71
Provided Data – IBM Connections on Intranet

- People graph (110,000,000 edges between 534,000 employees)
  - Connections of IBM employees through their interactions in IBM Connections, e.g.
    - X commented on Y’s blog
    - X and Y are members of the same community
    - X and Y co-edited a wiki

- Raw Entity Data
  - For all IBM Connections applications provide all publicly created entities and their meta data
    - Entities (id, date, related people (author, commenter, sharer, etc.), parent community)
    - Employee profiles (anonymized id, friends, manager, division, country)
    - 534,000 user profiles, 26,000 communities, 40,000 blog entries, 16,000 Wiki elements, 21,000 shared files and 10,000 forum threads
CAT dashboard

Undesirable Role Decomposition

From Date: [Calendar]
To Date: [Calendar]

Undesirable Role Decomposition

Graph showing changes in undesirable role decomposition over time.
Observe Activity in community

Determine the level of activity (drop) in the community - check for off-topic discussions; unanswered threads.

- Observe Activity in community
- Simulate thread ordering...
- Evaluate simulation result...
- Apply policy changes
Policy Forum Restriction simulator

It can be simulated how the limitation of new threads in specific forums (e.g., the number of new threads in Forum #50 is limited to 10 per day) affects the time to a quality reply to a thread in all forums. The idea is that the attention of users is steered to other forums if the activity in a specific forum is restricted.