THE NET OF THE 21st CENTURY: Concepts across the Interspace

Bruce Schatz
CANIS Laboratory
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
www.canis.uiuc.edu, schatz@uiuc.edu

CNI Spring Meeting
April 26, 1999  Washington, DC
THE THIRD WAVE OF NET EVOLUTION

ARPA\r\nINTERNET\r\nINTERSPACE

PACKETS
OBJECTS
CONCEPTS

CONCEPT SPACES

- from Objects to Concepts
- from Syntax to Semantics
- Infrastructure is Interaction with Abstraction

Internet is packet transmission across computers
Interspace is concept navigation across repositories
LEVELS OF INDEXES

Technology

Engineering

Electrical

IEEE

FORMAL (manual)

INFORMAL (automatic)

communities

groups

individuals
THE DISTRIBUTED WORLD

- Community Repositories in the Interspace
- Every Person performs Every Role

- USER request
- LIBRARIAN reference
- INDEXER classify
- PUBLISHER quality
- AUTHOR generate
CONCEPTS ACROSS THE INTERSPACE

INDEXING

NAVIGATING

SWITCHING

FLYING

CANIS Community Architectures for Network Information Systems
www.canis.uiuc.edu
SCALABLE SEMANTICS

- Automatic indexing
- Domain-Independent indexing
- Statistical clustering

- Compute Context of
  - concepts within documents
  - documents within repositories
CROSS-OVERS IN SEMANTIC INDEXING

COLLECTIONS

1960 - laboratory

2000 community

discipline

2010 -

COMPUTERS

Watch Personal Desktop Department Shared Supercomputer GRID
COMPUTING CONCEPTS

‘92: 4,000 (molecular biology)
‘93: 40,000 (molecular biology)
‘95: 400,000 (electrical engineering)
‘96: 4,000,000 (engineering)
‘98: 40,000,000 (medicine)
SIMULATING A NEW WORLD

- *Obtain discipline-scale collection*
  - MEDLINE from NLM, 10M bibliographic abstracts
  - human classification: Medical Subject Headings
- *Partition discipline into Community Repositories*
  - 4 core terms per abstract for MeSH classification
  - 32K nodes with core terms (classification tree)
- *Community is all abstracts classified by core term*
  - 40M abstracts containing 280M concepts
  - concept spaces took 2 days on NCSA Origin 2000
- *Simulating World of Medical Communities*
  - 10K repositories with > 1K abstracts (1K w/ > 10K)
COMMUNITY PROCESSING

CATEGORY MAPS

CONCEPT SPACES

document context within repository

phrase context within document

CONCEPTS

noun-phrase extraction

DOCUMENTS

COMMUNITY PROCESSING

INDEXING

CANIS Community Architectures for Network Information Systems

www.canis.uiuc.edu
INTERSPACE NAVIGATION

- Semantic Indexes for Community Repositories

- Navigating Abstractions within Repository
  - concept space
  - category map

- Interactive browsing by Community experts
Navigation in MEDSPACE

For a patient with Rheumatoid Arthritis

- Find a drug that reduces the pain (analgesic)
- but does not cause stomach (gastrointestinal) bleeding
Concept Search

InterSpace Remote Access: Search and Navigation

Domains

[+] [+] Artificial Intelligence
[+] [+] Asian Economics
[+] [+] Colon Cancer
[+] [+] Osteo Arthritis
[+] [+] Rheumatoid Arthritis

Search

bleeding

Submit Search

51 cr gastrointestinal bleeding on volunteer
anal bleeding
bleeding diathesis
ingestion bleeding
ingestion bleeding incidence
getting process
getting salicylate
getting bleeding
cerebral bleeding
complication with spontaneous bleeding
evidence of gastrointestinal bleeding
getting bleeding
gastrointestinal bleeding in patient
gastrointestinal bleeding under treatment
gastrointestinal bleeding
gastrointestinal bleeding among patient
gastrointestinal bleeding
hæmorrhæmis rectal bleeding
hypoprothrombinemic bleeding
intestinal bleeding
lower gastrointestinal bleeding
moderate bleeding
occult bleeding
possible gastrointestinal bleeding
reverse intraoperative bleeding

www.canis.uiuc.edu
Concept Navigation

- bursal bleeding
- cerebral bleeding
- complication with spontaneous bleeding
- evidence of gastrointestinal bleeding
- gastric bleeding
- gastrointestinal bleeding in patient
- gastrointestinal bleeding under treatment
- nonsteroidal anti-inflammatory drug
- drug
- kirchheiner b
- ameliorating effect
- trang l
- simple analgesic
- maintenance therapy
  - [doc] Maintenance therapy of rheumatoid arthritis and of osteoarthritis
  - [doc] Nonsteroidal anti-inflammatory drugs, arthritis, and gastritis
  - [doc] Selective in vivo removal of rheumatoid factor by an extracorporeal method
  - [doc] Sulfasalazine in the treatment of rheumatoid arthritis. A randomized, double-blind, placebo-controlled study
  - [doc] Innovative treatment approaches for rheumatoid arthritis
  - clark b
  - anti-inflammatory agent
  - mason dl
  - gastrointestinal bleeding
  - polisson rp
  - rennie ja
  - random sample
  - tenidap
  - paracetamol
Maintenance therapy of rheumatoid arthritis and of osteoarthrosis with proglumetacin.

Bozsoky S., Zahumenszky Z.

Maintenance therapy with proglumetacin was studied in an open investigation in 25 patients with classical or definite rheumatoid arthritis and in 34 patients with osteoarthrosis. Proglumetacin (150 mg) was administered twice daily, at meals, and therapy was continued without interruption for 12 months. Patients with rheumatoid arthritis showed a progressive improvement in objective parameters (erythrocyte sedimentation rate, haemoglobin) and in semi-objective parameters (pain, articular tenderness, number of painful joints, morning stiffness, grip strength, Ritchie articular index). Patients with osteoarthrosis also showed a progressive improvement in objective parameters (movement angles of the affected joints) and in subjective parameters (pain, mobility, response to therapy). During the study, 4 patients reported occasional nausea (3 with vomiting), and 3 reported episodes of slight headache. These symptoms did not require interruption of treatment. Haematology, blood chemistry and urinalysis were not adversely affected by the treatment. Five drop-outs were
Maintenance therapy of rheumatoid arthritis and of osteoarthritis with proglumetacin.

Bozsoky S., Zahumenszky Z

Maintenance therapy with proglumetacin was studied in an open investigation in 25 patients with classical or definite rheumatoid arthritis and in 34 patients with osteoarthritis. Proglumetacin (150 mg) was administered twice daily, at meals, and therapy was continued without interruption for 12 months. Patients with rheumatoid arthritis showed a progressive improvement in objective parameters (erythrocyte sedimentation rate, haemoglobin) and in semi-objective parameters (pain, articular tenderness, number of painful joints, morning stiffness, grip strength, Ritchie articular index). Patients with osteoarthritis also showed a progressive improvement in objective parameters (movement angles of the affected joints) and in subjective parameters (pain, mobility, response to therapy). During the study, 4 patients reported occasional nausea (3 with vomiting), and 3 reported episodes of slight headache. These symptoms did not require interruption of treatment. Haematology, blood chemistry and urinalysis were not adversely affected by the treatment. Five drop-outs were

affected joint
articular index
bozsoky s
definite rheumatoid arthritis
drop-out
intolerance
maintenance therapy
mg
objective parameter
orthopaedic surgery
osteoahtrosis
osteoahtrosis group
patient with osteoarthritis
patient with rheumatoid arthritis
proglumetacin

rheumatoid arthritis
Proglumetacin, 450 mg/day, was administered during 1 month to 32 rheumatoid arthritis patients in anatomical and functional stage II or III. Overall, more than 50% of pts responded to the treatment, in particular those in stage II, in whom Ritchie’s articular index, nightly and daily pain significantly decreased. Those pts in stage III who did not respond to proglumetacin even when administered at high dose, did not respond to indomethacin, either. Tolerance resulted very good, as only 3 complaints of mild and transient epigastric pain and 1 of skin rash were observed, but none related to the C.N.S. High efficacy and good tolerance set therefore proglumetacin aside from indomethacin, A.S.A. and most of the recent non steroidal anti-inflammatory agents.
Nonsteroidal anti-inflammatory drugs, arthritis, and gastrointestinal bleeding in elderly in-patients.

In a study of 551 consecutive admissions to an acute geriatric admitting unit in England in a four-month period, 33 cases of upper gastrointestinal (UGI) bleeding were diagnosed. There was a close relationship between UGI bleeding and the use of nonsteroidal anti-inflammatory drugs (NSAIDs), and it is argued that this is a causal link. In view of the dangers of NSAID ingestion by elderly patients, many of whom are using them for analgesic rather than anti-inflammatory reasons, it is suggested that simple analgesics such as paracetamol should be used for maintenance therapy of osteoarthritis in the elderly, while NSAIDs are used for treatment of true arthritis in selected patients.
Nonsteroidal anti-inflammatory drugs, arthritis, and gastrointestinal bleeding in elderly in-patients.

In a study of 531 consecutive admissions to an acute geriatric admitting unit in England in a four-month period, 33 cases of upper gastrointestinal (UGI) bleeding were diagnosed. There was a close relationship between UGI bleeding and the use of nonsteroidal anti-inflammatory drugs (NSAIDs), and it is argued that this is a causal link. In view of the dangers of NSAID ingestion by elderly patients, many of whom are using them for analgesic rather than anti-inflammatory reasons, it is suggested that simple analgesics such as paracetamol should be used for maintenance therapy of osteoarthritis in the elderly, with NSAIDs reserved for treatment of true arthritis in selected patients.
SWITCHING

In the Interspace...

- each Community maintains its own repository
- Switching is navigating Across repositories
- use your vocabulary to search another specialty
ENGINEERING SESSION

[Image of a computer screen showing a software engineering interface with categories, concept spaces, and search results.]
Engineering Categories & Concepts

CATEGORIES

660 Automotive Engineering
670 Naval Architecture and Marine Engineering
680 Railroad Engineering
690 Materials Handling
700 Electrical Engineering, General
710 Electronics and Communication Engineering
720 Computers and Data Processing
721 Computer Circuits and Logic Elements
721.1 Computer Theory (Includes Formal Logic, Automata)
721.2 Logic Elements
721.3 Computer Circuits
722 Computer Hardware
722.1 Data Storage (Equipment and Techniques)
722.2 Computer Peripheral Equipment
722.3 Data Communication (Equipment and Techniques)
722.4 Digital Computers and Systems
722.5 Analog and Hybrid Computers
723 Computer Software, Data Handling and Applications
723.1 Computer Programming
723.1.1 Computer Programming Languages
723.1.2 COM-1 Design Patterns
723.2 Database Systems
723.3 Artificial Intelligence

CONCEPT SPACES

C6110B SOFTWARE ENGINEERING TECHNIQUES

object database

software object database
software environment database
environment database
object-oriented database

software configuration management

object model
configuration management
software engineering environment
constraint maintenance
specifying complex
handling change

Switch Using: 🔹 Term only
🔹 With Related Terms

Search: object database

Word Order: 🔶 As Given
🔹 Any

Interspace Navigator
Further Concept Navigation

C6110B SOFTWARE ENGINEERING TECHNIQUES

object database
- software object database
- software environment database
- environment database
- object-oriented database
- yang, y.
- specify software
- knowledge-based software engineering
- complex object

software configuration management
- object model
- configuration management
- software engineering environment
- constraint maintenance
- specifying complex
- handling change

C6110B SOFTWARE ENGINEERING TECHNIQUES

software configuration management
- configuration management system
- configuration management plan
- configuration management discipline
- ieee standard dictionary
- software engineering standard
- ieee guide
- produce reliable software
- configuration management issues
- std
- ansi/ieee std
- ieee std

management plan
- revision control system

Switch Using:
- Term only
- With Related Terms

Search:
- object database

Word Order:
- As Given
- Any
Searching via Concept Suggestion

Search for: revision control system

Complex objects in the Doris system
Rossi, K. | Yairi, R.

The Doris system provides methods and tools for software configuration management on top of the RCS (revision control system) version control system. The Doris system supports hierarchical and complex software objects called configurations, and the concept is applicable for both software development and documentation. Some of the conceptual and technical features of the Doris system are discussed and implementation on the Unix file system is presented [configuration management, software engineering].
Switching Across Repositories

<table>
<thead>
<tr>
<th>Complex Object</th>
<th>Complex Object</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database Management System</td>
<td>Query Language</td>
</tr>
<tr>
<td>Object Version</td>
<td>Multidatabase System</td>
</tr>
<tr>
<td>Configuration Management</td>
<td>Heterogeneous Database System</td>
</tr>
<tr>
<td>Saake, G.</td>
<td>Component Objects</td>
</tr>
<tr>
<td>Multiple View</td>
<td>Deductive Database</td>
</tr>
<tr>
<td>Object-Oriented Paradigm</td>
<td>System Supporting</td>
</tr>
<tr>
<td>Information System</td>
<td>Abstract Data Types</td>
</tr>
<tr>
<td>Office Information System</td>
<td>Data Structure</td>
</tr>
<tr>
<td>Object Database</td>
<td>Multimedia Database</td>
</tr>
<tr>
<td>Object Aggregation</td>
<td>Specification Language</td>
</tr>
<tr>
<td>Relational Data Model</td>
<td>Object Model</td>
</tr>
<tr>
<td>Ehrich, H.</td>
<td>Database System</td>
</tr>
<tr>
<td>Design Application</td>
<td>Formal Specification Language</td>
</tr>
</tbody>
</table>

Switch Using: Term only

Search: Object Database

Word Order: As Given

C An I S
Community Architectures for Network Information Systems
www.canis.uiuc.edu
CONCEPT SWITCHING

- “Concept” versus “Term”
  - set of “semantically” equivalent terms
- Concept switching
  - region to region (set to set) match

Semantic region

Concept Space

term

Concept Space
Medicine Session

Regulation by CDF/LIF and retinoic acid of multiple ChAT mRNAs produced from distinct promoters.

The cholinergic differentiation factor/neuromedin B inhibitory factor (CDF/LIF) and retinoic acid (RA) induce in sympathetic neurons, a switch from the norepinephrine to the cholinergic neurotransmitter phenotype. In particular, these molecules alter the activities of the biosynthetic enzymes choline acetyltransferase (ChAT), tyrosine hydroxylase (TH) and dopamine beta-hydroxylase (DBH). Recently, five rat ChAT mRNA species have been identified although no data have yet been reported concerning their production and regulation in sympathetic neurons. By use of the reverse transcription-polymerase chain reaction technique, we analyzed the effects of CDF/LIF and RA on the levels of ChAT, TH and DBH mRNAs.
Categories and Concepts

Domains
- Colonic Neoplasms
- Colonic Polyps
- Adenomatous Polyposis Coli
- Colorectal Neoplasms
- Sigmoid Neoplasms
- Genes, Regulator
  - Genes, araC
  - Genes, nef
  - Genes, Nitrogen Fixation
  - Genes, pX
  - Genes, rev
  - Genes, Switch
  - Genes, tat
  - Genes, vif
  - Genes, vpr
  - Genes, vpu
- Regulatory Sequences, Nucleic Acid

Search: Colorectal Neoplasms, Hereditary Nonpolyposis
- families with hereditary nonpolyposis colorectal cancer
- hereditary cancer
- hereditary colon cancer
- hereditary colorectal cancer
- hereditary non-polyposis colon cancer
- Hereditary non-polyposis colorectal cancer
- Hereditary non-polyposis colorectal cancer
- Hereditary nonpolyposis colon cancer
- Hereditary nonpolyposis colorectal cancer
- nonpolyposis colorectal cancer
- DNA mismatch repair genes
- families with hereditary nonpolyposis colorectal cancer
- Support, U.S. Gov't, P.H.S.

mismatch repair genes
- DNA mismatch repair genes
- repair genes
- Hereditary nonpolyposis colorectal cancer
- human mismatch repair genes
- nonpolyposis colorectal cancer
- mismatch repair gene
- Support, U.S. Gov't, P.H.S.
- DNA mismatch repair gene
- Polymerase Chain Reaction
- Polymorphism, Single-Stranded Conformati
Concept Switching

CANIS Community Architectures for Network Information Systems

www.canis.uiuc.edu
Document Retrieval

Regulation by CDF/LIF and retinoic acid of multiple ChAT mRNAs produced from distinct promoters.

The cholinergic differentiation factor/leukaemia inhibitory factor (CDF/LIF) and retinoic acid (RA) induce in sympathetic neurones, a switch from the noradrenergic to the cholinergic neurotransmitter phenotype. In particular, these molecules alter the activities of the biosynthetic enzymes choline acetyltransferase (ChAT), tyrosine hydroxylase (TH) and dopamine beta-hydroxylase (DBH). Recently, five rat ChAT mRNA species have been identified although no data have yet been reported concerning their production and regulation in sympathetic neurones. By use of the reverse transcription polymerase chain reaction technique we analysed the effects of CDF/LIF and RA on the levels of ChAT, TH and DBH mRNAs.
• **Landscape as category map visualization**

• **Valleys are semantic clusters**

• **Hills are semantic distances**

• **Traversal across multiple levels of abstraction**
THE NET OF THE 21st CENTURY

- **Beyond Objects to Concepts**
- **Beyond Search to Analysis**
- **Problem Solving via Cross-Correlating Multimedia Information across the Net**

- *Every community has its own special library*
- *Every community does semantic indexing*

- *The Interspace is true Cyberspace*
The Zen of the Net