Superimposed Information

Collaborators

- **Concept & Architecture:** Lois Delcambre, Sun Murthy
- **Corporate Information Systems:** Lougie Anderson, Radhika Reddy
- **Clinical Information:** Paul Gorman, Joan Ash, Mary Lavelle, Jason Lyman
- **Forest Management:** Tim Tolle
- **Tobacco Documents:** Jim Pankow, Patty Toccalino
- **SLIM & SLIMPad:** Shawn Bowers, Longxing Deng, Mat Weaver
- **SIDEWALK (NSDL):** Lillian Cassel, Ed Fox, Eric Hanson, David Archer, Adam Lowry, Susan Price, James Terwilliger

What is Superimposed Information?

Information added over a base layer of documents and data to aid in re-finding, reuse, explanation, analysis, repurposing, reorganization.

What's the interest for IR?

- Helping with tasks after the information has been found
- Related to Personal Information Management (PIM), such as re-finding information.

Superimposed Information has Existed for Millennia

- Concordances
- Religious and legal commentaries
- Interlinear text
- Card catalogs
- Annotated bibliographies

Concordance

Index of occurrences of a term or phrase in a corpus, often with context

Much Ado about Nothing [I, 3] 372 ***Don John***
Will it serve for any *model* to build mischief on?

Pericles [II, 2] 759 ***Simonides***
It's fit it should be so; for princes are
A *model* which heaven makes like to itself:
As jewels lose their glory if neglected,
So princes their renowns if not respected.

Hamlet [V, 2] 3701 ***Hamlet***
I had my father's signet in my purse,
Which was the *model* of that Danish seal;

DOMESTIC
d., kind, and submissive to their Tit 2.05
DOMINEERING
not as d. over those in your charge 1Pe 5.03
DOMINION
let them have d. over the fish of Gen 1.26
and have d. over the fish of the 1.28
are you indeed to have d. over us? 37.08
By Jacob shall d. be exercised, and Num 24.19
the Philistines had d. over Israel. Ju 14.04

Commentary

Explanation or analysis of a text, often piece by piece, sometimes with excerpts

CHAPTER 2

Ge 2:1. THE NARRATIVE OF THE SIX DAYS' CREATION CONTINUED. The course of the narrative is improperly broken by the division of the chapter.

1. the heavens--the firmament or atmosphere.

host--a multitude, a numerous array, usually connected in Scripture with heaven only, but here with the earth also, meaning all that they contain.

were finished--brought to completion. No permanent change has ever since been made in the course of the world, no new species of animals been formed, no law of nature repealed or added to. They could have been finished in a moment as well as in six days, but the work of creation was gradual for the instruction of man, as well, perhaps, as of higher creatures (Job 38:7).

Ge 2:2-7. THE FIRST SABBATH.

2. and he rested on the seventh day--not to repose from exhaustion with labor (see Isa 40:28), but ceased from working, an example equivalent to a command that we also should cease from labor of every kind.

Legal Commentary

Generally refers to laws or previous cases

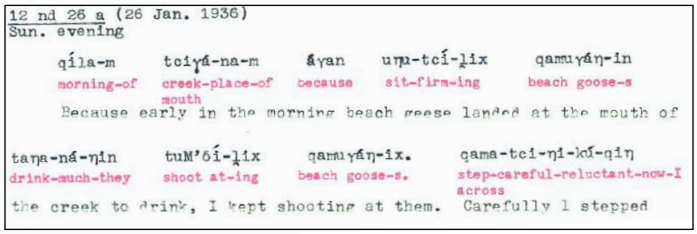
Although § 66.0501(2), Stats., requires only that a governing body member resign before being appointed in order to be eligible to hold an office or position the selection to which is vested in the governing body and which was not created during the member's term, a governing body member should resign prior to applying to ensure that he or she does not violate §§ 946.13(1)(a) and 19.59(1)(a) Stats. 10/29/01.

Note reports U.S. Supreme Court decision to address how much protection the Fifth Amendment's public use requirement provides for individuals whose property is being condemned for the purpose of economic development in the eminent domain case from the Connecticut Supreme Court, *Kelo v. City of New London*, 268 Conn. 1, 843 A.2d 500 (Ct. 2004) and two eminent domain decisions from state courts also addressing the "public use" requirement, *Bailey v. City of Mesa*, 206 Ariz. 224, 76 P.3d 898 (App. 2003) and C*ounty of Wayne v. Hathcock*, 2004 WL 1724875, ___ Mich. __, 684 NW.2d 765 (2004). 9/30/04.

Interlinear Text

Annotation interspersed with text.

12) Now when Jesus heard that John had been arrested, He left [the province of Judea] **and went to Galilee. 13)** [After] **leaving Nazareth** [i.e., a major town in Galilee]**, He came to live in Capernaum,** [a city] **beside Lake Galilee, near the districts of Zebulun and Naphtali. 14)** [This was] **so that what Isaiah the prophet had spoken would be fulfilled, saying, 15) The districts of Zebulun and Naphtali, near the lake, beyond** [i.e., where it enters] **the Jordan in Galilee,** [the territory] **of the Gentiles; 16) the people who lived** [there] **in** [spiritual] **darkness saw a great light. Light shone upon these people, who lived in an area that was under the shadow of death.**



Addressing

Notice that some of these varieties of information make use of a standard addressing scheme on the base layer.

- Book, Chapter, Verse
- Play, Act, Scene, Line
- Section, Subsection, Paragraph, Subparagraph

Less Formal SI: Annotating*

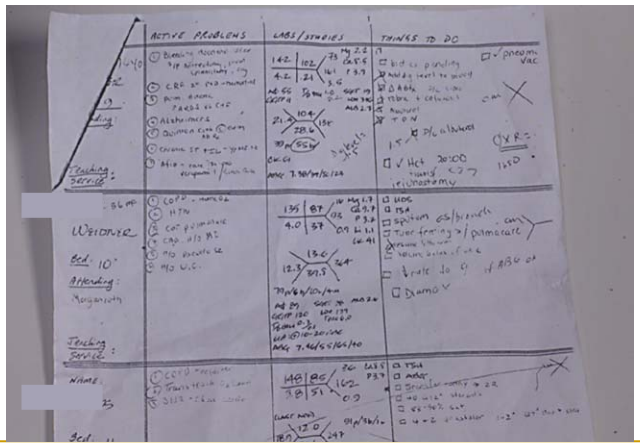
What's going on here?

- Highlighting
- Commenting
- Classifying
- Grouping
- Linking

*Picture courtesy of Prof. James Pankow, Dept. of EBS, OGI

A Bundle of Excerpts: Is This SI?

Certainly makes selections, adds structure



SI Has Carried over to Digital World

Bookmarks or Favorites in web browser

- A personal collection of locations in the World-Wide Web
- Simple interface: add & browse

XLinks between web locations

Comments feature in document applications

- Add, move, delete, check-off
- Sort, summarize

Such as notes in Acrobat

Superimposed Applications

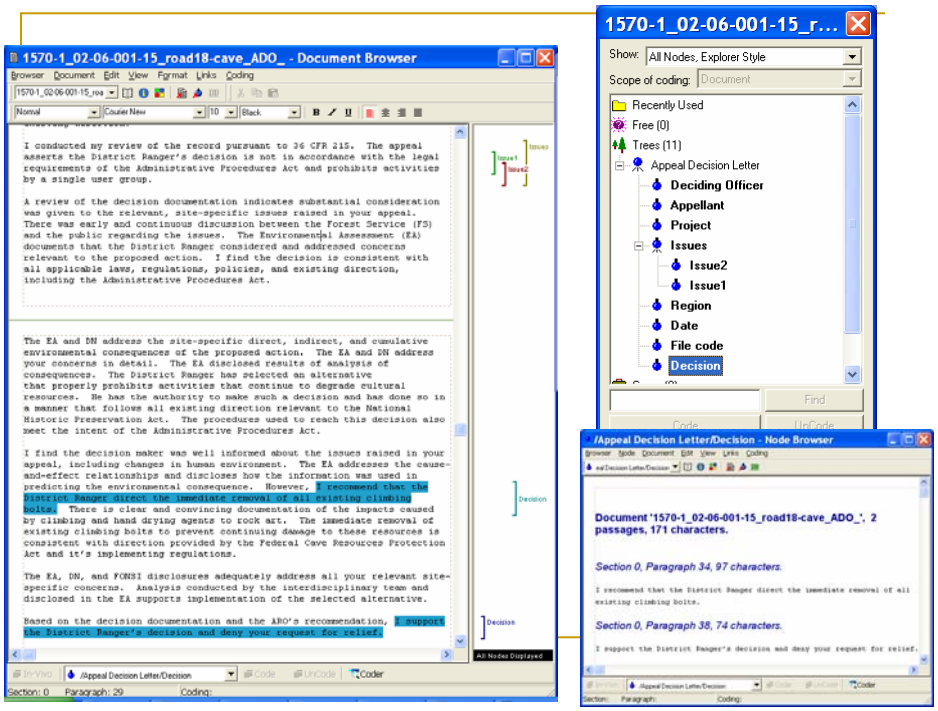
Software tools that make use of SI

Can be a separate interface or extend a base application

NVivo

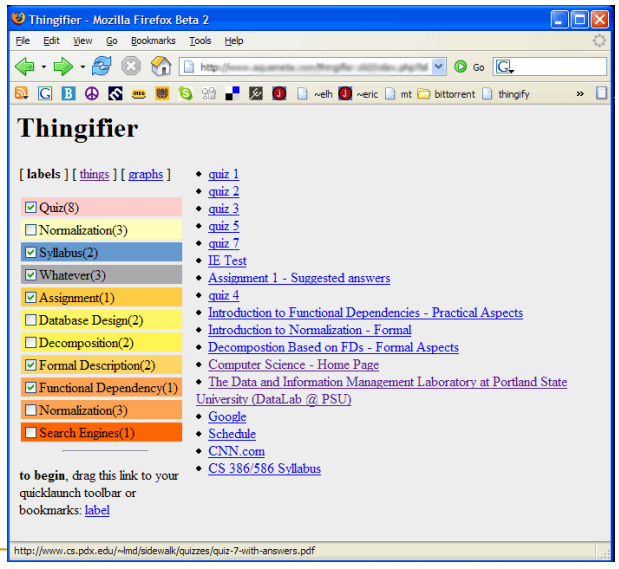
Coding for qualitative analysis

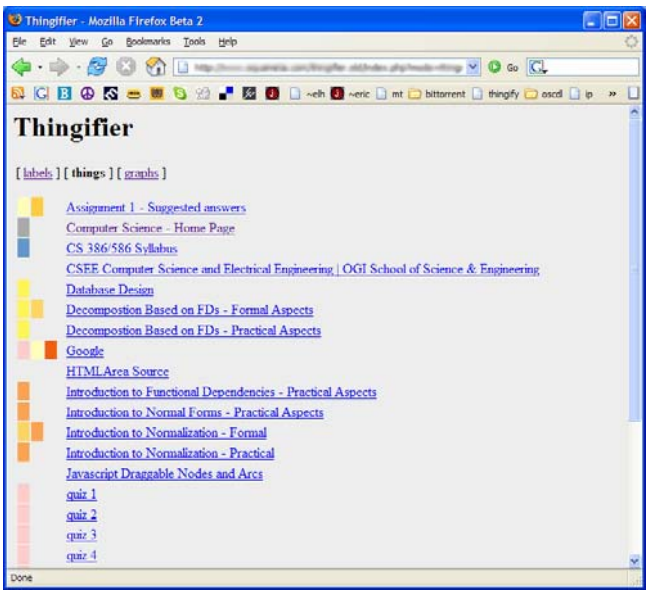
- *Nodes* associated with one or more text regions, can be organized in *trees* and *cases*
- Can add *attributes* to nodes and documents
- Annotate text with *DataBites*
- Can *link* out to segments of multimedia
- Create arbitrary *sets* of nodes and documents



Some Superimposed Applications We've Built

Thingifier:
Attach labels to "things" on the web

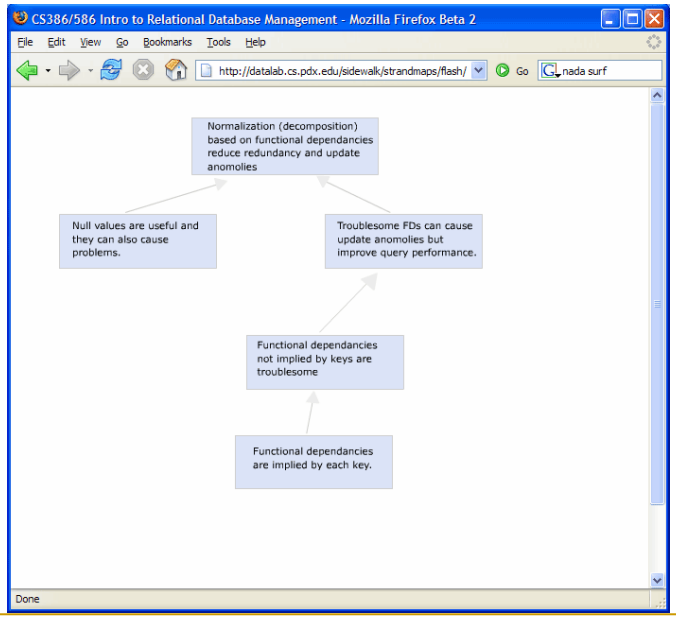


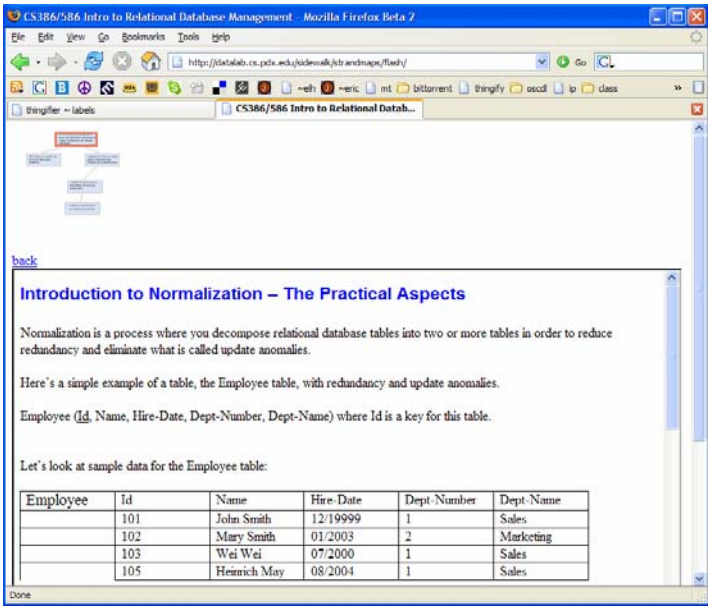


Strand Map Browser

Show dependencies among topics in a body of knowledge

- We've added connections to related materials
- Using to organize course units


CS386/586 Intro to Relational Database Management - Mozilla Firefox Beta 2
File Edit View Go Bookmarks Tools Help
http://datalab.cs.pdx.edu/sidewalk/strandmaps/flash/
Go
nada surf
Normalization (decomposition) based on functional dependancies reduce redundancy and update anomolies
Null values are useful and they can also cause problems.
Troublesome FDs can cause update anomolies but improve query performance.
Functional dependancies not implied by keys are troublesome
Functional dependancies are implied by each key.
Done


CS386/586 Intro to Relational Database Management - Mozilla Firefox Beta 2
back
Introduction to Normalization – The Practical Aspects
Normalization is a process where you decompose relational database tables into two or more tables in order to reduce redundancy and eliminate what is called update anomalies.
Here's a simple example of a table, the Employee table, with redundancy and update anomalies.
Employee (Id, Name, Hire-Date, Dept-Number, Dept-Name) where Id is a key for this table.
Let's look at sample data for the Employee table:
Employee
Id
Name
Hire-Date
Dept-Number
Dept-Name
101
John Smith
12/19999
1
Sales
102
Mary Smith
01/2003
2
Marketing
103
Wei Wei
07/2000
1
Sales
105
Heinrich May
08/2004
1
Sales
Done

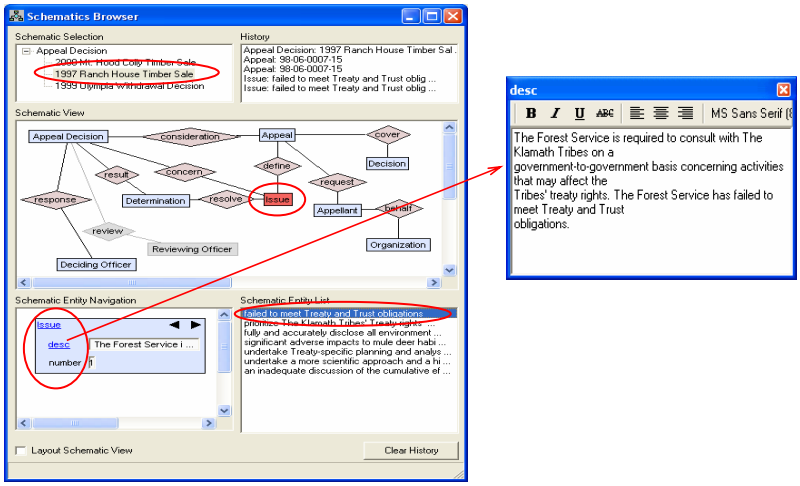
Superimposed Schematics Browser*

Use ER model to identify and browse document collection for particular domain

A kind of "superware" – can be customized to different domains

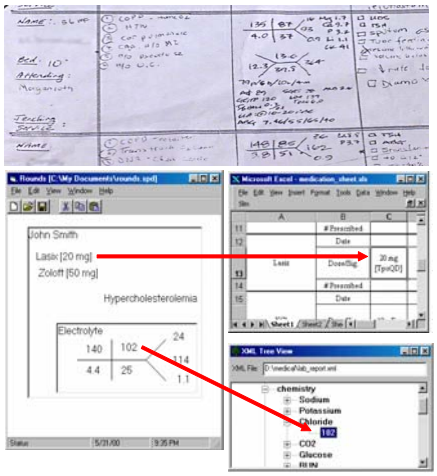
Example on US Forest Service appeals process

*Developed originally by Shawn Bowers



SLIMPad

Trying to create a digital version of a clinical bundle



Sidepad Demo

Motivated by "bundle building" behavior

Successor to SLIMPad and RIDPad

Simple information model

- Items with labels that mark into a base documents
- Groups of items and groups
- Labeled links (also with marks)

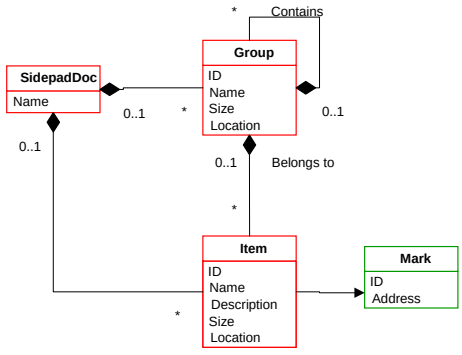
Also support for accessing content and context of a marked region

Superimposed Information Model

Superimposed model can be different from base layer (which might be relatively unstructured)

- XLink: labeled graph (or hypergraph)
- Acrobat notes: location-comment pairs
- NVivo: nodes w/ multiple marks, w/ attributes, organized in trees and sets
- Thingifier: labeled collections (for classification)
- Strand Map Browser: Directed acyclic graphs (DAGs)
- Schematics Browser: Instance of ER schema with a mark on each entity or relationship instance (and a `Mark` value type)

Sidepad Superimposed Model



Stand-off vs. In-line Approaches

- Modify base layer versus reference it
- Issues with Stand-off
  - Coping with changes in base layer
  - Traversing from base to superimposed layer
- Issues with In-line
  - May not own base layer
  - Alternative SI on same base document
  - How do you have SI span multiple documents?

Granularity of Addressing

What are identifiable elements in base layer?

- Bookmark list: whole web pages (or anchor points provided by content)
- Sidepad: regions selectable by base application
- XLink: Whatever XPath can point to
- Interlinear text: Span of what an annotation refers to isn't always clear

Embedded vs. Encapsulated Marks

- CARTE had mark-description info embedded directly in superimposed information

  (CARTE was our first superimposed application, adapted from Topic Maps)
  - Hard to share marks across different SI models
- Sidepad has a repository of marks
  - SI models use handles to them

Semantics of Marks

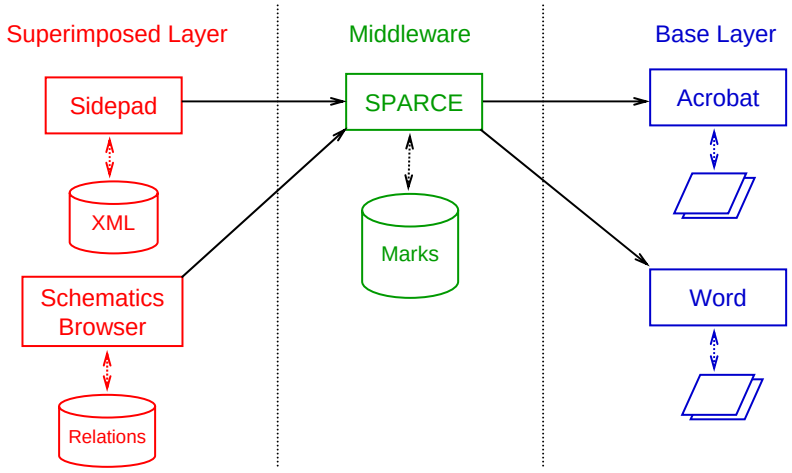
- Who interprets marks
  - Superimposed application
  - Base application
  - External agent (XPath in XLink)
- What can you ask about marks?
  - Identity, Equivalence
  - Containment, Overlap
  - What's in it (content)
  - What's around it (context)

Our Approach: SPARCE

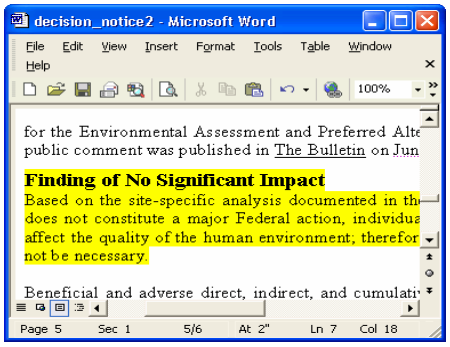
Superimposed Pluggable Architecture for Context and Excerpts

- Support multiple SI models
- Stand-off SI
- Address base layer at subdocument granularity
  - Address space defined by base application
  - Marks created in those applications
- Encapsulated marks, interpreted by base applications
  - Reactivate, Content, Context

SPARCE Architecture Overview



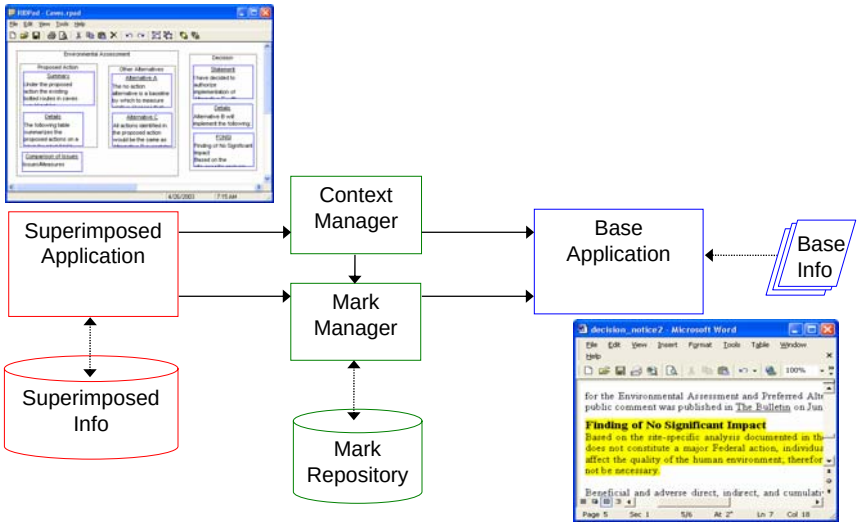
Excerpts and Contexts



Name	Value
Excerpt	Finding of No Significant Impact … not be necessary
Font name	Times New Roman
HTML	Finding of No Significant Impact Based on …

- *Excerpt* is the content of a marked region
- *Context* is information related to a mark
- A mediator called *context agent* retrieves content & context of a mark

SPARCE Reference Model



Different Approaches to Marks

Are marks held in a common repository, or distributed among different superimposed applications?

- Common Repository
  - Provides sharing across different superimposed applications and superimposed document instances
  - Provides collections for querying
  - But binds superimposed document to a repository

"Sparcels"

"Self-carrying code" for marks

`sparce:markfodder=SUVBZ2VudH…kJ`

Encodes application, document, region, user

- Still need access to mark manager, context manager
- Need to be able to register a handler for the `sparce` namespace
- Potentially portable, but embeds addresses of documents

API to Marks

Can be through standard PL bindings, for example COM objects from Visual Basic or C++

Can provide services to activate base application, extract content, context, (update?)

Also can be through a URI

- Can provide different services via "query" part
- `sparce://datalab.cs.pdx.edu/markid=WordMark2003Apr24071826TYEEsmurthy?action=goto`

Where is Interpretation Done?

- Local server

  Requires software to be present locally
- Remote "rendering" service

  Needs to reformat base document to show highlighting

  Can't invoke local base application
- Remote "script" service

  Returns a script to do local activation

  Tends to be browser specific

Bi-Level Queries

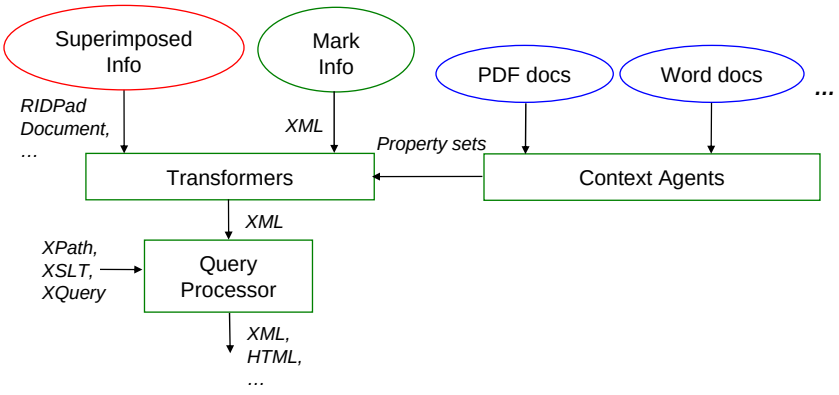
Ask queries over the superimposed information plus the content and context of marked regions

- Locate marked information based on its connections
- Select and format information for display
- Use superimposed information as "lever" to do database style transformation and restructuring on non-DB information

Build a common logical model for query

- Have talked about dual-language queries

Basic Bi-level Query System



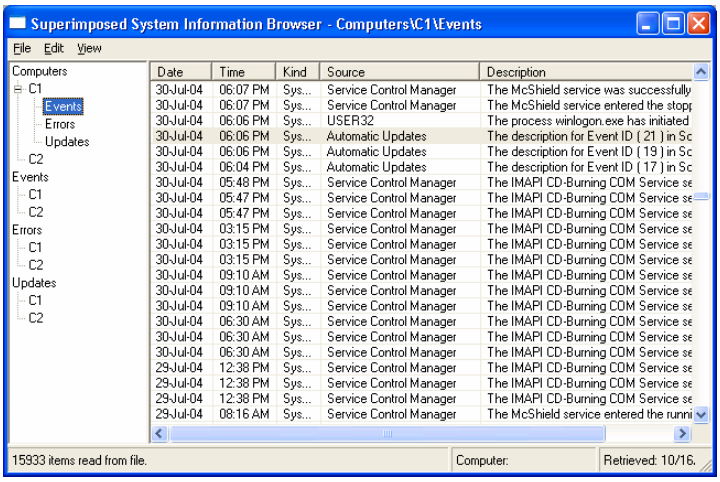
What's Neat about Bi-Level Queries

- Only mediate what we use (unlike approaches that mediate entire sources)
- Works with different structures of superimposed information
- Useful in transforming and repurposing documents
- Adding a little bit of structure lets us do a lot of querying

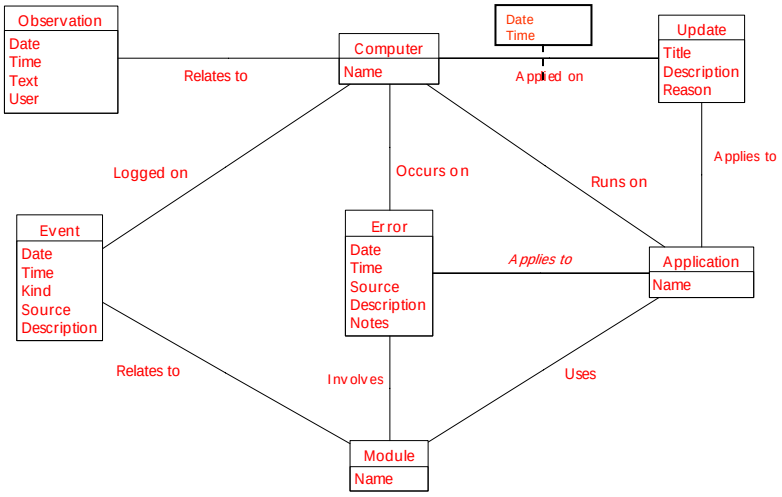
The PC-Admin Browser

- Allows a system (network) administrator to browse information about computers in a network
  - Applications installed and the modules they use
  - Updates applied
  - Errors recorded
  - Logged application, system, and security events
  - User comments

PC-Admin Mockup



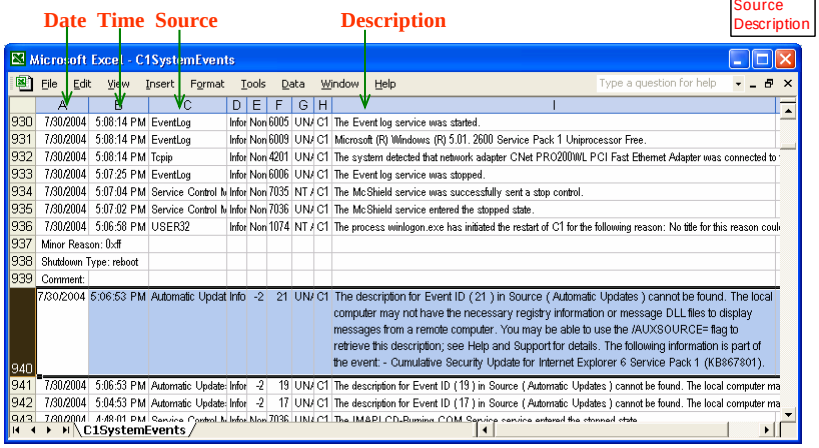
A Conceptual Schema*



* All entities have key attribute ID (not shown); all relationships are many-many

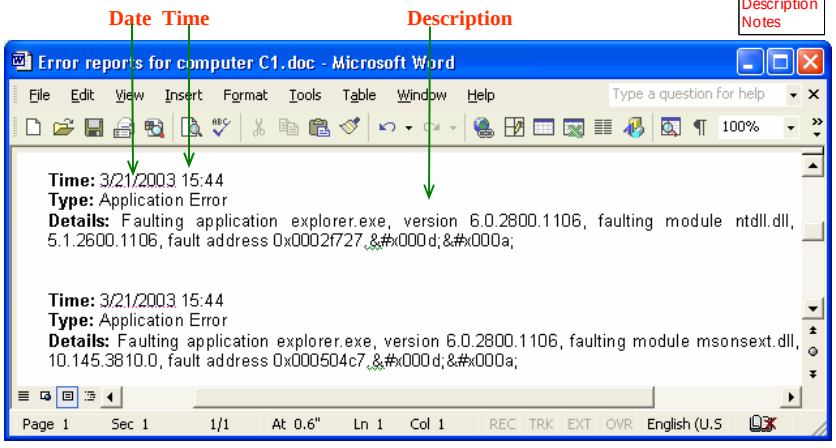
Base Layer: Event Log

Event
Date
Time
Kind
Source
Description



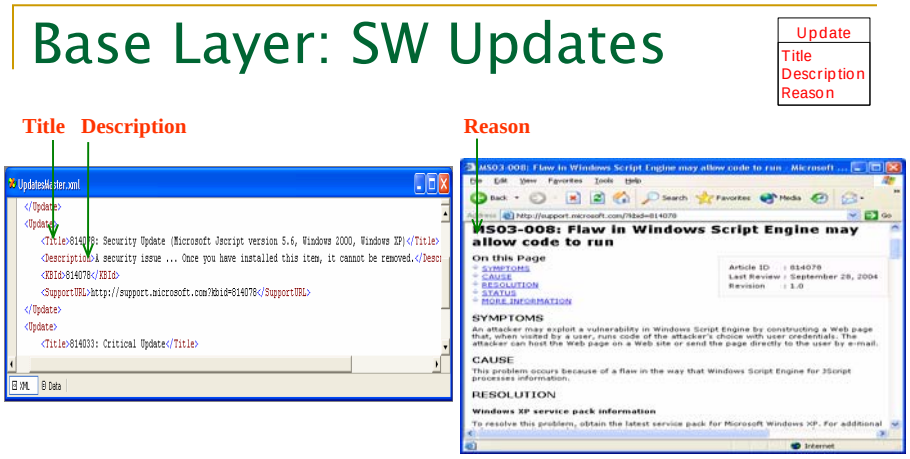
Base Layer: Error Reports

Error
Date
Time
Source
Description
Notes



Uniform structure, but mapping is not clean: Date and Time are both in Time field

Base Layer: SW Updates



Data is heterogeneous and distributed: some data in XML, some in HTML

Structure varies: Support URL not always defined, HTML page structure varies widely

Example Query

- Retrieve all errors MS Word caused in the last week

```
SELECT * FROM Error
WHERE EDate BETWEEN CURRENT_DATE AND
  CURRENT_DATE - INTERVAL '6' DAY
AND Description LIKE '%Word.exe%'
```

- `Error` is a view: the attribute `description` is retrieved from the base layer when the view definition is executed

Another Example Query

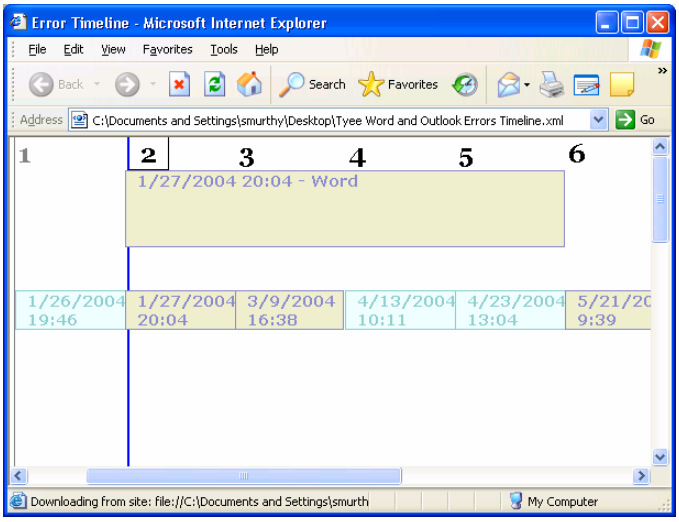
Create a timeline of errors related to MS Word and MS Outlook

```
SELECT EDate, ETime, Description
FROM Error
WHERE Description LIKE '%word.exe%'
OR Description LIKE '%Outlook.exe%'
```


Sample Results

EDate	ETime	Description
1/26/2004	19:46	Hanging app…Outlook.EXE…
1/27/2004	20:04	Faulting app…winword.exe…
3/9/2004	16:38	Hanging app…winword.EXE
4/13/2004	10:11	Faulting app…Outlook.EXE…
4/23/2004	13:04	Hanging app…Outlook.EXE…
5/21/2004	9:39	Faulting app…winword.exe…
5/26/2004	14:05	Faulting app…winword.exe…

Result as Timeline*



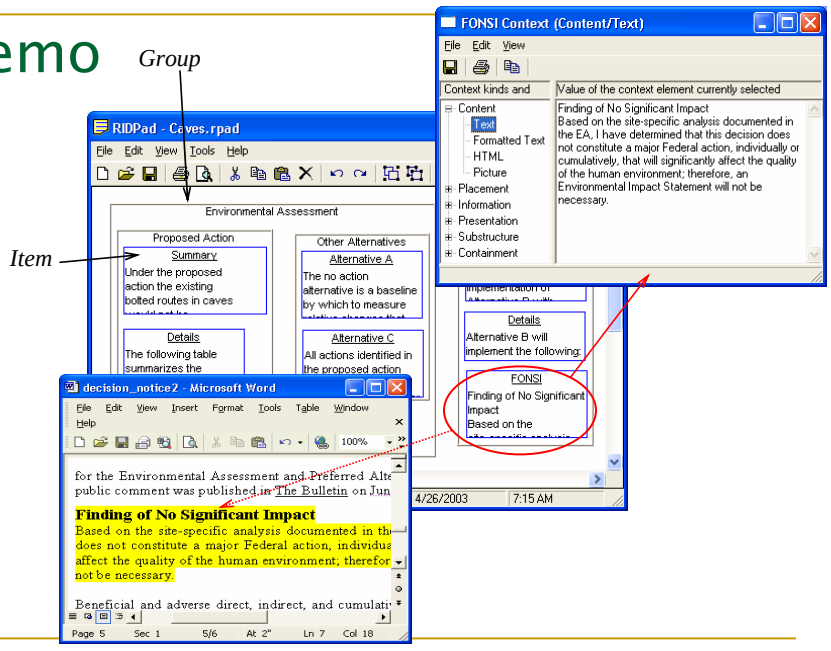
*Drawn using an XML transformation based on work of Nicolas Kruchten. Timeline is non-linear

Summary

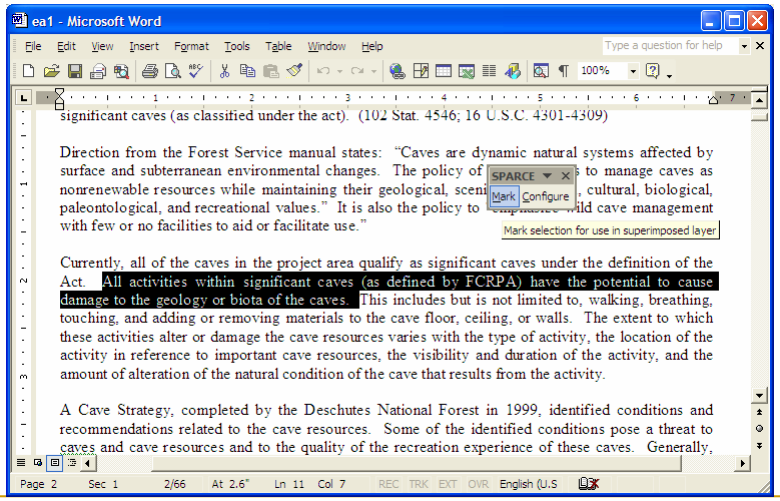
- Superimposed information has been around almost as long as writing
- Wealth of digital information presents even more opportunities for SI
  - Added structure
  - Information bridging
  - Virtual documents
- Interesting modeling issues
  - What's a mark?
  - Adapting a data model for SI
  - Moving between SI models

Back-up Slides

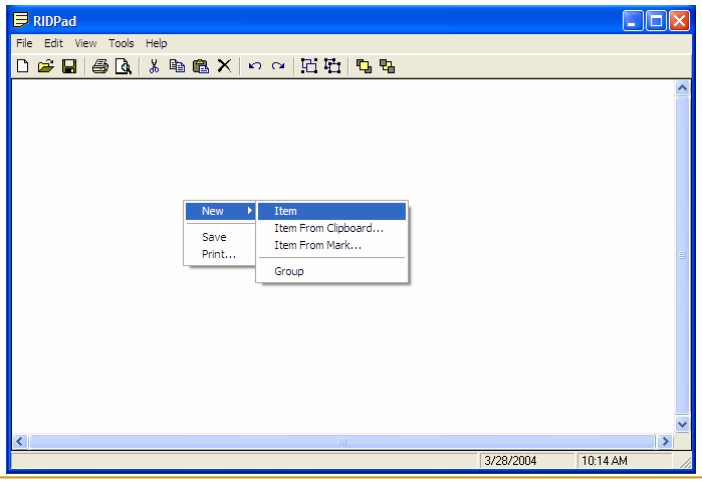
Demo



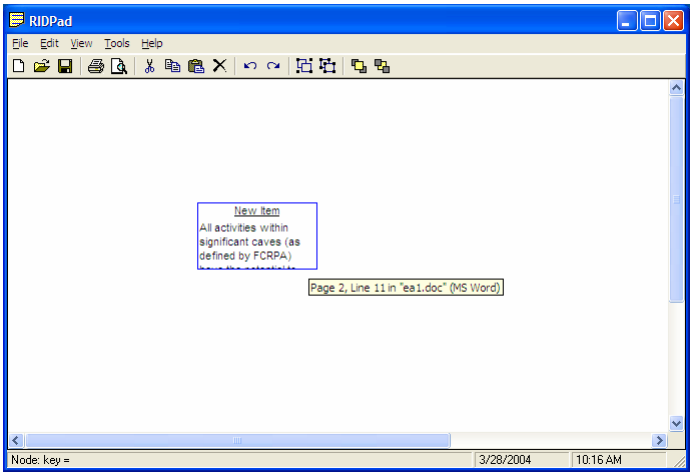
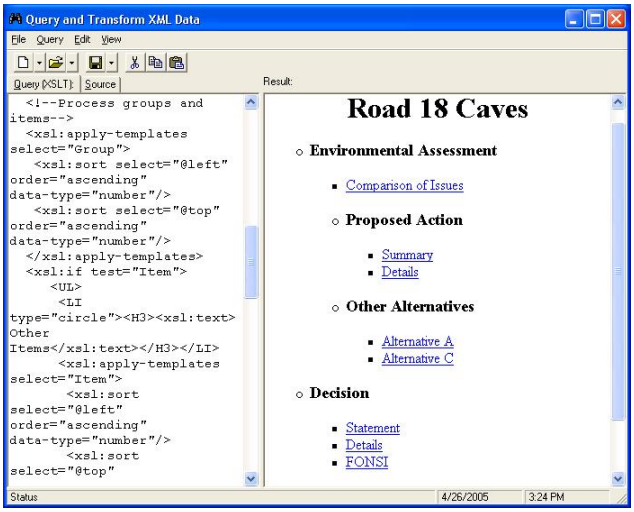
Creating a Mark



Creating a Mark



Creating a Mark



Transformation via Query