Structure and navigation for electronic publishing

John Tillinghast
Giordano Beretta

Hewlett-Packard Company

Problem

- Drowning in data, starving for information
- Current approach: search engines
  - search for keywords
  - search for concepts
  - search for similar images
- Most of the result is irrelevant to the information sought
- How can we find the needle in the haystack?
  - paint needles in red
  - make needles radioactive
  - use a real big magnet
  - store needles separately from hay

Simple solution?

- World Wide Web authoring tools do not assist organizing the contents
- The World Wide Web is a hypertext
- The implementors of CAI systems (Plato, XS-0) had learned that a system should be trees
  - rooted
  - no cycles
  - exactly one path between any two nodes
- Trees are too regimented — each item should not have to be just on one branch
- Graphs do not allow to extend the number of relationships — human relationships are too complicated

Example

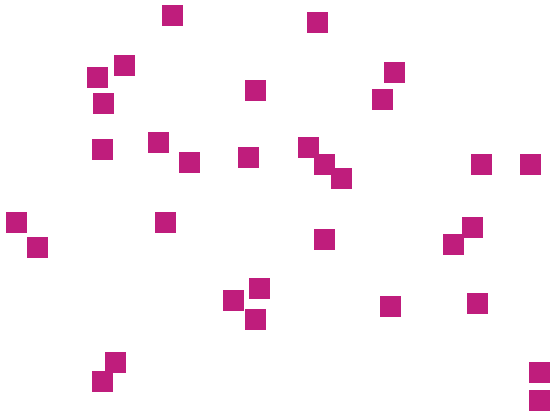
- A Web-based wedding album
- A relation based on kinship
  - there are certain characters — two families, guest, officiator, …
  - characters have different histories and interactions
- A relation based on the time-line
  - arrival of guests
  - small-talk
  - meeting new people
  - ceremony
  - reception
- Other relations: residence, transportation used, …

How did hypertext start?

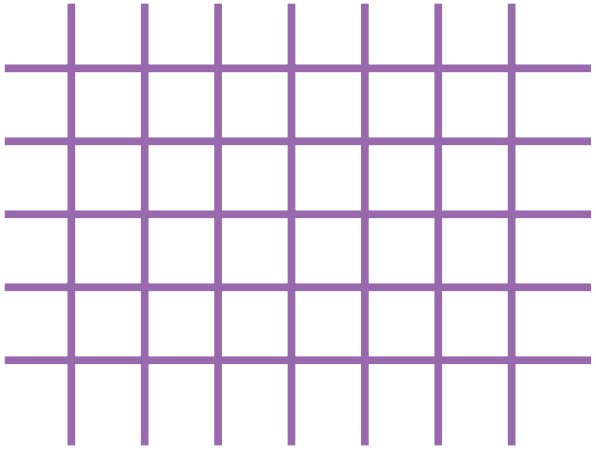
- Vannevar Bush (Memex) & Doug Engelbart (Augment)

- Doug Engelbart: categorization of evolving knowledge
- Recorded dialog
  - diaries, notes, address books, captions, databases
- External intelligence
  - guide books, ontologies
- Knowledge products
  - scrap books, Web photo albums

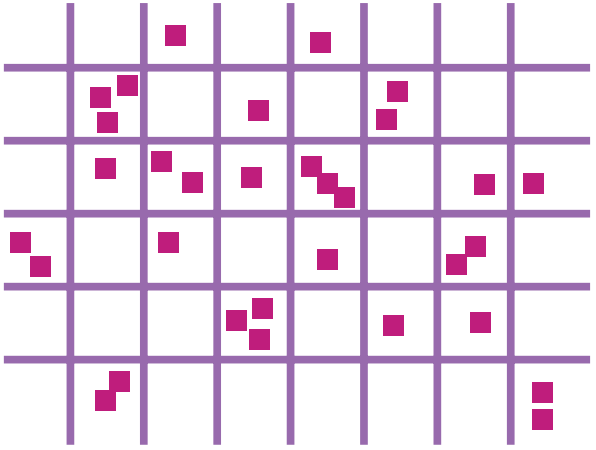
Categorization of evolving knowledge



information
recorded dialog



structure
external intelligence



knowledge

General Nicolas Bourbaki

Requirement 1: need a tool to easily define structures

- New interpretation of mathematics after 1935
- Relational construct: a set with relations
- System of axioms represents properties of constructs
- Mathematical creativity: find new constructs by defining maps that preserve the relations
- Two-step approach
  - find a good system of axioms
  - find a good isomorphic construct

Definitions — abstract structure

- *Abstract structure:* equivalence class of isomorphic relational constructs
- *Relational construct:* set with n-tuple of relations
- *System of axioms:* properties of relational constructs
- *Structure:* relations for a given system of axioms for a construct
- Example: group structure — (G, *) is a group
  - G0: G a set, * a binary relation on G
  - G1: * associative
  - G2: there is neutral element e
  - G3: each $x \in G$ has an inverse

Ariadne's thread

- Humans communicate events by narrating stories
- For Leibniz, the author finds a linear order in a set of information and creates knowledge in the form of a linear thread

The true method must provide us with a filum Ariadnes, that is to say a kind of sensitive and coarse means that guides the mind, in the same way as lines drawn in geometry and the type of operations that are prescribed to apprentices in Arithmetic. Without that our mind would not know how to go along a long path without straying.

Gottfried Wilhelm Leibniz

Requirement 2: to support storytelling, the tool must provide a canonical linearization

Software tool opportunity

- Mathematically: lattices
- Algorithms that manage structures and can traverse them

Requirement 3: need a software infrastructure

- Hard design problem: define the best partition between human and tool systems
- Human system is good at
  - categorization (find correct class)
  - associations (story telling, Ariadne's thread)
- Tool system
  - enabling technology: data structures, ADT, OOPS
  - needed: editors for the systematic definition of structures

Ontologies

- Ontology: explicit specification of a conceptualization
- A formal and declarative representation which includes the vocabulary (or names) for referring to the terms in that subject area and the logical statements that describe what the terms are, how they *are* related to each other, and how they *can or cannot* be related to each other

- A good ontology can be reused in similar situations
- Ontologies scale well — e.g., SHOE used to compile an ontology of infectious diseases involving hundreds of thousands of terms

Wedding album example

- Example of an ontology:
  - http://www-ksl.stanford.edu/
  - go to Ontolingua Server
  - load the ontology Wedding-Pics
  - see paper for further instructions
- Once the ontology has been specified, the ontology editor with its inference engine is no longer necessary
- A specific very simple application can be written that allows to author and read a set of Ariadne's threads based on a recorded dialog

Custom browser option

- Ontolingua frames can be translated into Java classes
- Java classes can be encapsulated into Beans
- A visual editor can be used to rapidly build a browser as a Java applet that represents some specific knowledge
- Ideally suited of simple vertical applications

Extensible Markup Language (XML)

- Data format for structured doc. interchange on the Web
  - a metalanguage to let authors design their own markup language
  - an abbreviated version of SGML tailored to Web applications
- More powerful than HTML
  - extensibility: can define new tags and attribute names
  - structure: document structure can be nested (tree instead of list)
  - appearance: XSL (eXtensible Style Language)
  - validation: a grammar can be supplied for structural validation
- Applications of XML
  - information discovery can be tailored to individual users
  - different views of the same data can be presented to different users
  - processing load can be transferred from server to client
  - mediation between heterogeneous databases

XML in action

- A community of practice develops & agrees on a Document Type Definition (DTD)
- Information components are stored in a repository
- Server creates documents (with DTD) on the fly, client takes into account user preferences and viewing options
- Embedded Java applet makes document active
- Microsoft's XML parser for Internet Explorer 4.0
  - http://www.microsoft.com/standards/xml/xmlparse.htm
  - other structured document specialists: Adobe, ArborText, Chrystal, DataWare, Documentum, Folio, Inforium, Inso, OpenText, PIT, Texcel, Vignette, …

Use of the ontology

- Current ontologies are somewhat tedious for direct use

- Aids in the normalization of the database
- Basis for a front-end to the database
- Instead of searching for a keyword, can search also for larger or smaller keywords
- More powerful enumeration than just “sort by”

Conclusions

- Evolution of e-publishing

Decade	Buzz word	Example	Key technology
70s	WYSIWYG	Bravo, Word	GUI
80s	object based, OLE	Tioga, FrameMaker	API, foundation classes
90s	Web, dynamic hypermedia	ODB + XML + Java	structure definition tools

- We have found that ontology editors are a viable tool for defining structures

Technology is the knack of so arranging the world that we do not have to experience it

Max Frisch