ONTOPEDIA
The Identity of Everything

Everything is a Subject

The vision of subject-centric computing

Steve Pepper, Ontopedia
pepper.steve@gmail.com

Topic Maps 2008 – Oslo, Norway 2008-04-04

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Everything is miscellaneous

- ☑ **Icebergs**
- ☒ Eleanor Rosch
- ☑ **Bush**
- ☑ **Lot**
- ☒ The forms of the clouds in the southern sky on the morning of April 30, 1882
- ☑ **Hamlet (?)**
- ☑ **Sisu**
- ☑ **Fuzzzy**
- ☑ **Copernicus**
- ☑ **Semantic Web**
- ☒ Russian numerals
- ☑ **Aristotle**
- ☑ **Wittgenstein**
- ☒ The feathers of spray lifted by an oar on the Río Negro on the eve of the Battle of Quebracho
- ☑ **OO programming**
- ☒ Ireneo Funes
- ☑ **Steve Pepper**

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Vannevar Bush and Hypertext

Vannevar Bush
1945
As We May Think
Memex

Doug Engelbart
1962
Augmenting Human Intellect
NLS / AUGMENT

Bill Atkinson
1987
HyperCard

Ted Nelson
1965
"Hypertext"
Xanadu

Tim Berners-Lee
1989
Information Management: A Proposal
World Wide Web

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

"As We May Think"

- **Concerned with the problem of finding information**
  - Existing technology hopelessly out of date:
  - *The amount of information is being "expanded at a prodigious rate", but the means we use to find it is "the same as was used in the days of square-rigged ships"*
  - The solution is to get away from hierarchical systems of organization and adopt new techniques that reflect how the brain works

Vannevar Bush
1945
As We May Think
MEMEX

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Associative thinking

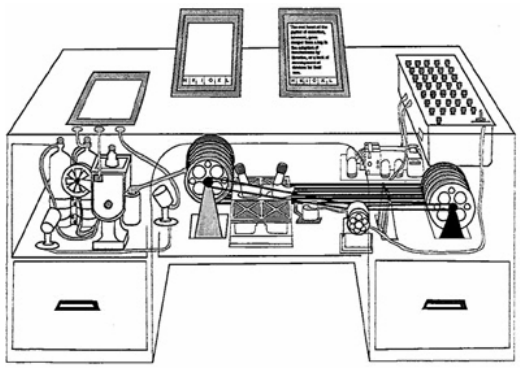
"The human mind … operates by association. With one item in its grasp, it snaps instantly to the next that is suggested by the association of thoughts, in accordance with some intricate web of trails carried by the cells of the brain… The speed of action, the intricacy of trails, the detail of mental pictures, is awe-inspiring beyond all else in nature."

Vannevar Bush: *As We May Think* (1945)

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Memex (memory extender)



A "sort of mechanized private file and library"

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

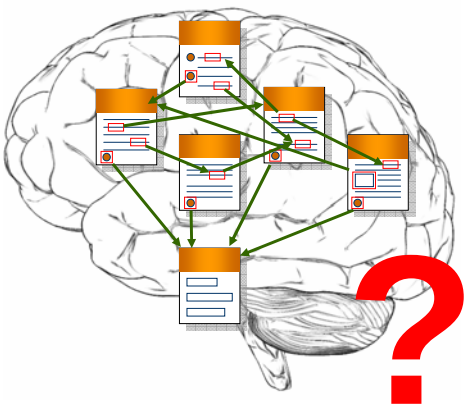
Memex (memory extender)

- **Consists of a desk containing**
  - a very large set of documents stored on microfilm
  - screens on which those documents are projected
  - a device for photographing new documents
  - a mechanism for retrieving documents at the push of a button
  - the ability to create links between documents
  - the ability to build trails through documents, add comments to documents, insert new documents, etc.
- **Note how everything revolves around documents**

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Is this how *you* think?

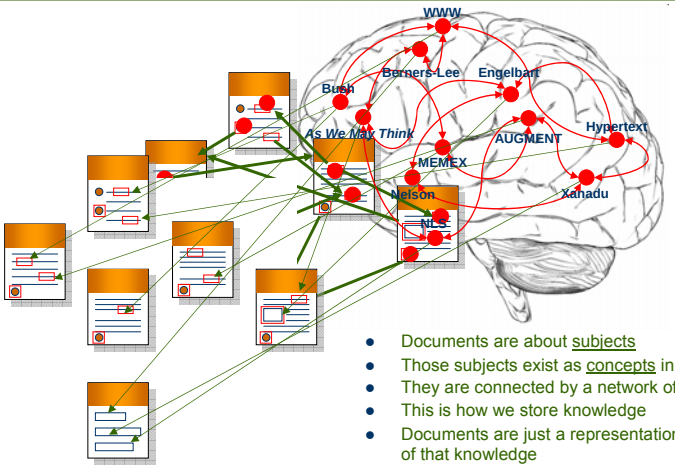


- **Is your head full of little documents all hyperlinked together?**
- ***I doubt it !***
- ***Mine certainly isn't !***
- **We don't think in terms of hyperlinked documents; we think in terms of concepts, and associations between concepts**

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

How we really think



- Documents are about subjects
- Those subjects exist as concepts in our brains
- They are connected by a network of associations
- This is how we store knowledge
- Documents are just a representation of some part of that knowledge

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Bush – right and wrong

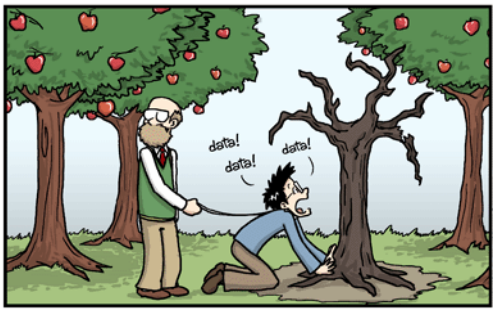
- Vannevar Bush was **right** that people think associatively
- He was **right** that organizing information in this way would make it easier to find
- But he was **wrong** in adopting a document-centric approach to the problem
- His basic idea – organize information "**as we may way think**" – was a great inspiration to Engelbart, Nelson, Atkinson, and Berners-Lee

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Barking up the wrong tree

- **But the Memex sent them all off in the wrong direction**
- **Hypertext has been barking up the wrong tree ever since**



- **And the Web, magnificent as it is, has made things "worse"**

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

"As We May Think" *(63 years on)*

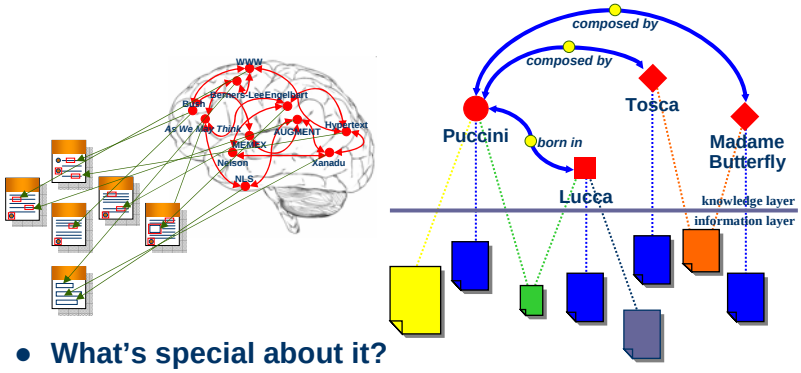
Vannevar Bush
1945
As We May Think
MEMEX

- **Concerned with the problem of finding information**
  - Existing technology hopelessly out of date:
  - *The amount of information is being "expanded at a prodigious rate", but the means we use to find it is "the same as was used in the days of ~~square-rigged ships~~ card catalogs"*
  - The solution is still to get away from hierarchical systems of organization and adopt new techniques that reflect how the brain works
  - That solution has to be subject-centric, not document-centric like the Web

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Which brings us to Topic Maps



- **What's special about it?**
  - #1 The TAO* model corresponds to how people think

* **T**opics + **A**ssociations + **O**ccurrences

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

TM as information architecture

- **<u>This</u> is what explains why TMs are ideal for web sites**
  - It really is computing "as we may think"
- **Subject-centric**
  - One page per topic (the concept of "subject page")
  - Page contents built primarily from names and occurrences
- **Associative**
  - Associations for navigating from one page (topic) to another
- **Example: topicmaps.com**

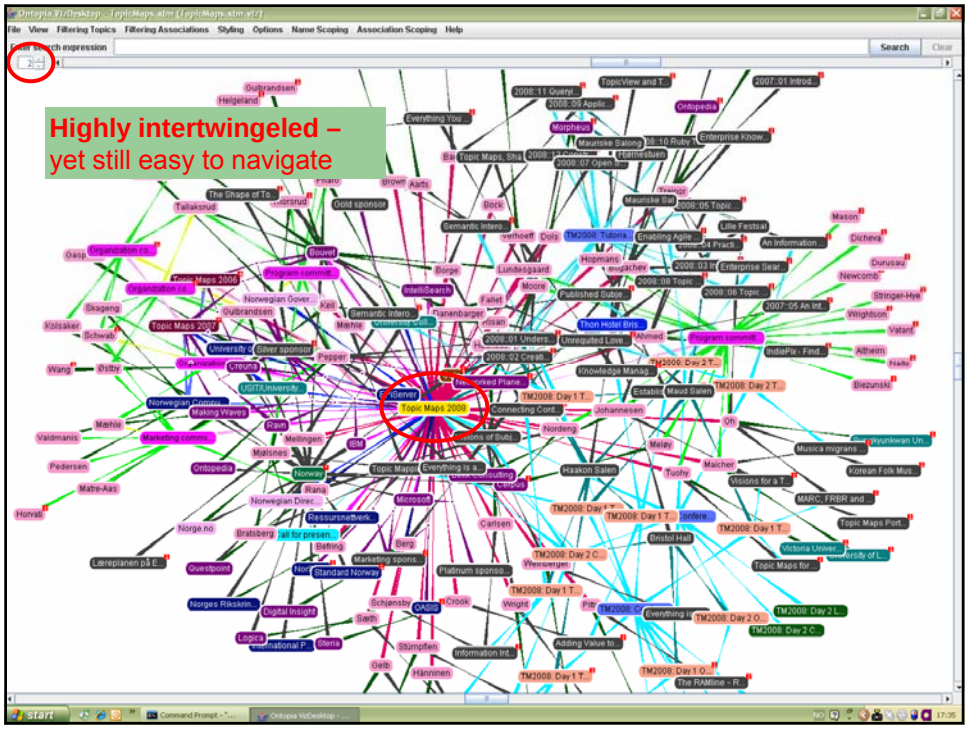
http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

topicmaps.com

100% topic map-driven

http://www.topicmaps.com/tm2008/pepper.ppt



ONTOPEDIA
The Identity of Everything

So is TM a portal technology?

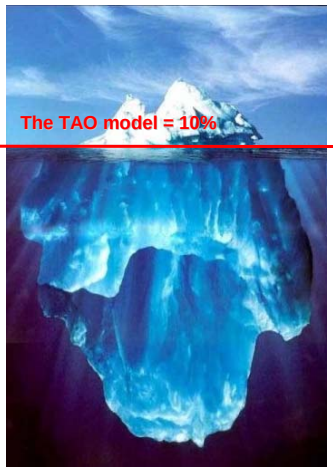
- **No, it's not**
- **Many people think so**
  - But it wasn't invented as such
  - It just turned out to be ideal for the purpose, because...
- **The underlying model is "as we may think"**
  - That model is subject-centric, not document-centric
- **Until recently most applications of Topic Maps were portals**
- **Now they are not, as this conference has shown**
  - (But the perception will persist, unless we all do something about it)

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

The tip of the iceberg

- **Today most applications use only the TAO model**
- **That means they use about 10% of the potential**
- **This is not a criticism**
  - Just something to be aware of lest you miss out on the major benefits
- **There's more to Topic Maps than the TAO**



http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

What else is there?

- **Scope**
- **Merging**

- **Generalized subject-centric computing**

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Scope: context is king

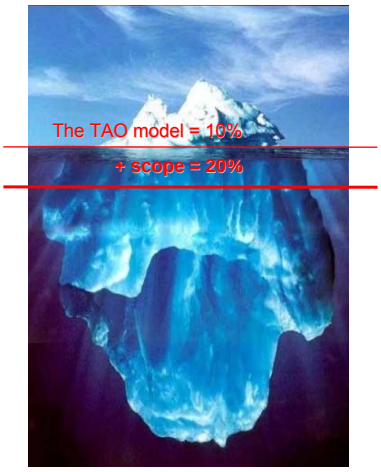
- **The TAO lets us express knowledge**
- **But knowledge has context**
  - Reality is ambiguous
  - Knowledge has a subjective dimension
  - Assertions may be valid in a one context but not another
- **Topic Maps has the concept of scope**
- **Scope enables the expression of contextual validity**
  - Permits multiple world views to coexist simultaneously
  - Allows us to handle the “miscellaneousness” of everything
- **Makes TM more than just a semantic technology**
  - It’s also a pragmatic technology
  - (Also in the sense that it’s ready to go today)

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Scope doubles the potential

- **Applications that use scope as well as the TAO can achieve 20% of the potential**
- **A Norwegian example:**
  - www.hoyre.no uses scope to enable over 400 different web sites (one per local branch) from a single topic map
- **The ability to merge topic maps more than doubles it again...**



http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Merging – global knowledge federation

- **Single most powerful feature**
- **Original motivation in 1991**
- **Business requirement**
  - Merge multiple, digital, back of book indexes in order to create a master index, without getting caught out by homonyms, synonyms, polysemes and the like
- **Merging has been there from day 1**
- **It's what enables global knowledge federation**
- **And it's why Lars Helgeland is wrong**

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

What's merging about?

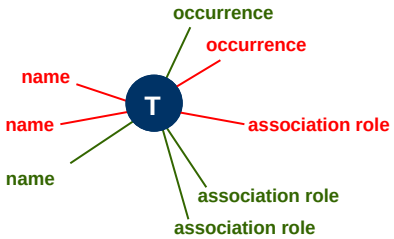
- **Topic Maps can be merged automatically**
  - Arbitrary topic maps can be merged into a single topic map
  - This cannot be done with databases or XML documents
- **Merging enables many advanced applications**
  - Information integration across repositories
  - Sharing and reusing taxonomies
  - Automated content aggregation
  - Distributed knowledge management
  - Global knowledge federation

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Principles of merging

- **By definition: Every topic represents exactly one subject**
- **Our goal: Every subject represented by just one topic**
  1. When two topic maps are merged, topics that represent the same subject should be merged to a single topic
  2. When two topics are merged, the resulting topic has the union of the characteristics of the two original topics



...and the resulting topic has the union of the original characteristics

(Demo of merging in the Omnigator?)

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

How can we achieve merging?

- **Well, we need to know when we – and our computers – are talking about the same thing**
- **Can't be done using names**
- **Almost every subject has multiple names**
- **For instance**
  - multiple languages
  - synonyms
  - polysemes
- **Name are notoriously unreliable for this stuff...**

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

And don't I know it!

pepper peper piper k'undo berbere pipor filfil فلفل bghbegh Պղպեղ jaluk biber бибэр piper пипер golmarich piper kani nayukon pebre hú-jiāo胡椒 pepř peber peper pepper peper pipro pipar pippuri poivre piper shitor pilpili პილპილი pfeffer piobar màsooroo pepa ipepile פלפל mari pipéri πιπέρι मिर्च mirch kua txob bors pipar merica pepe koshoo こしょう menasu ಮೆಣಸು burış Бұрыш mrech huchu후추 phik noi piper pipari pipirai mulagu lada povaair поваарь maricha fefer marich philphili pieprz kanu pimenta piper perets перец marica papar miris poper pepere pimienta pilipili peppar milagu மிளகு savyamu paminta phrík thai fowarilbu pepa biber perets перець mirch pilpel hạt tiêu pupur peprovník *u*phepha pepee pementa pebre peure pepre ペッパー

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

The exceptions are few

- **Mostly very specific and culture-dependent**
  - The Finnish word *sisu*
  - The Xhosa word *ubuntu*
- **Then there's the problem of homonyms**
  - Many names have multiple referents
  - Ubuntu, whatever its original meaning is also the name of a Linux distribution

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Consider *pepper*, if you will

- **Wikipedia's disambiguation page lists**
  - 13 different plants
  - 10 different people
  - 9 'other's
  - 3 'see also's

- **Norwegian adds another:**
  - ***gi pepper til noen:*** level criticism at someone

Microsoft-seieren får pepper

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Humans can tackle this

- **In natural language we get by using names**
  - Various strategies are used, including
  - Context
  - Negotiation
- **But computers aren't that smart**
- **How can they know when two symbols have the same referent?**
  - That is, when two topics represent the same subject
- **The only solution for computers is identifiers**

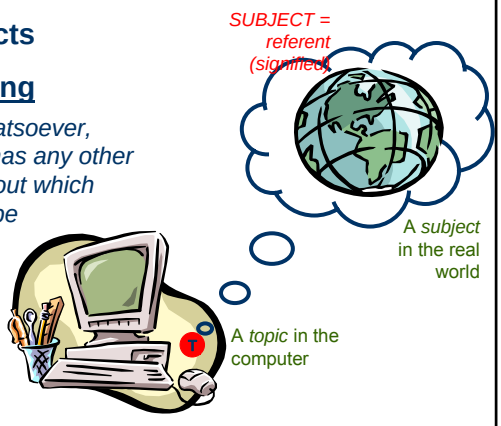
http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

The Topic Maps model of identity

The forms of the clouds in the southern sky on the morning of April 30, 1882

- **Topics represent subjects**
- **A subject can be anything**
  - *A subject is any "thing" whatsoever, whether or not it exists or has any other specific characteristics, about which anything whatsoever may be asserted by any means whatsoever."*
- **Everything is a subject**
  - as soon as a human has thought about it

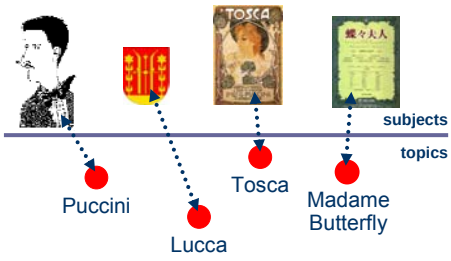


http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

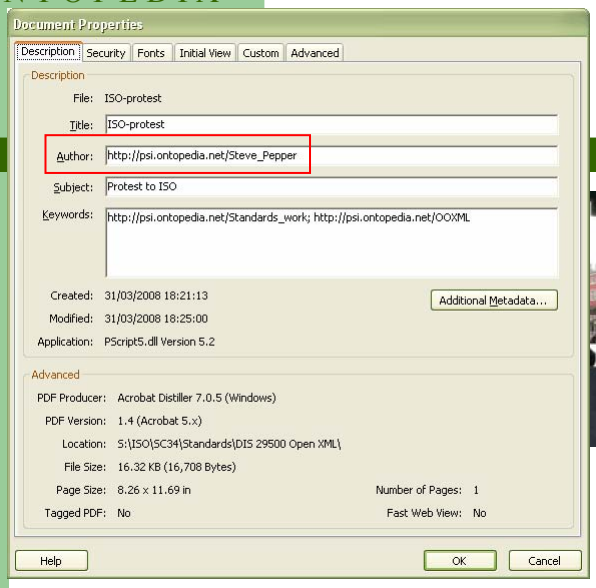
Subject identifiers

- **Meaning is expressed through the relationship between the representation and referent**
- **Aka intentionality:**
  - in topic maps, intentionality is captured using subject identifiers
  - makes it possible to know when two topics represent the same subject
  - allows topics to be shared across maps, and for maps to be merged



http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA



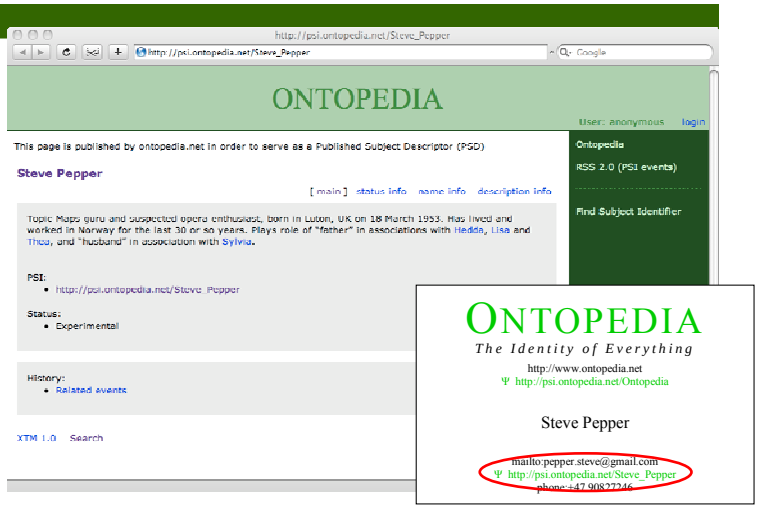
pepper får seieren

- **http://psi.ontopedia.net/Steve_Pepper**

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

A PSD for one Steve Pepper



http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Globally unique identifiers

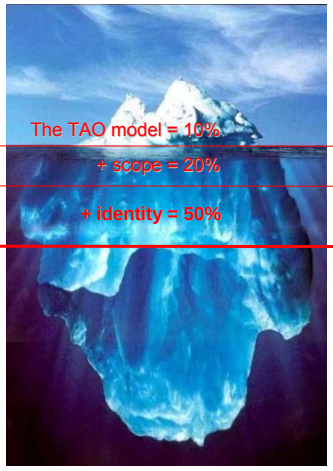
- **We're not the only ones thinking about this**
  - Librarians (I guess)
  - Publishers (ISBN, ISSN)
  - Document Object Identifiers (DOI)
  - Uniform Resource Names (URN)
- **Best current practice on the Web**
  - Use URIs
- **Emerging consensus is to use HTTP URIs**
- **The Topic Maps community has proposed a mechanism called Published Subjects**
  - It's time to get together and talk about this stuff

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Subject identifiers

- **PSIs are perhaps not the final answer**
  - But they're a pretty good stop gap
  - The potential more than doubles
- **But what about the other 50%?**
- **Learning from Web 2.0**
  - subject-centric tagging
  - subject-centric wikis
  - subject-centric blogging
- *(At this point, Pepper turns to the vendors present)*



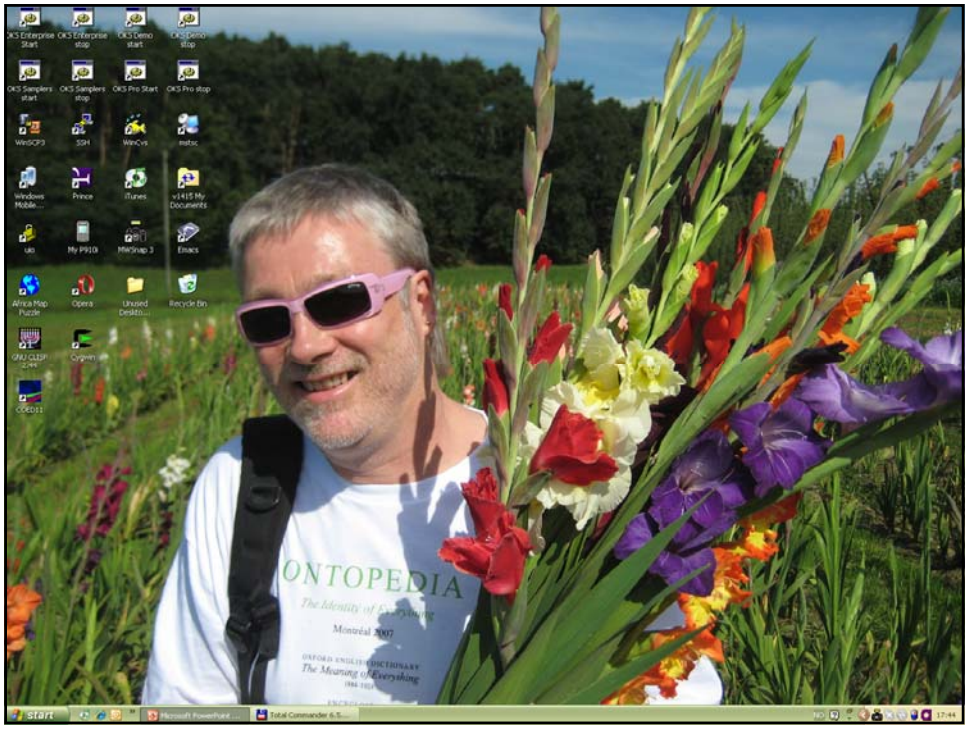
http://www.topicmaps.com/tm2008/pepper.ppt

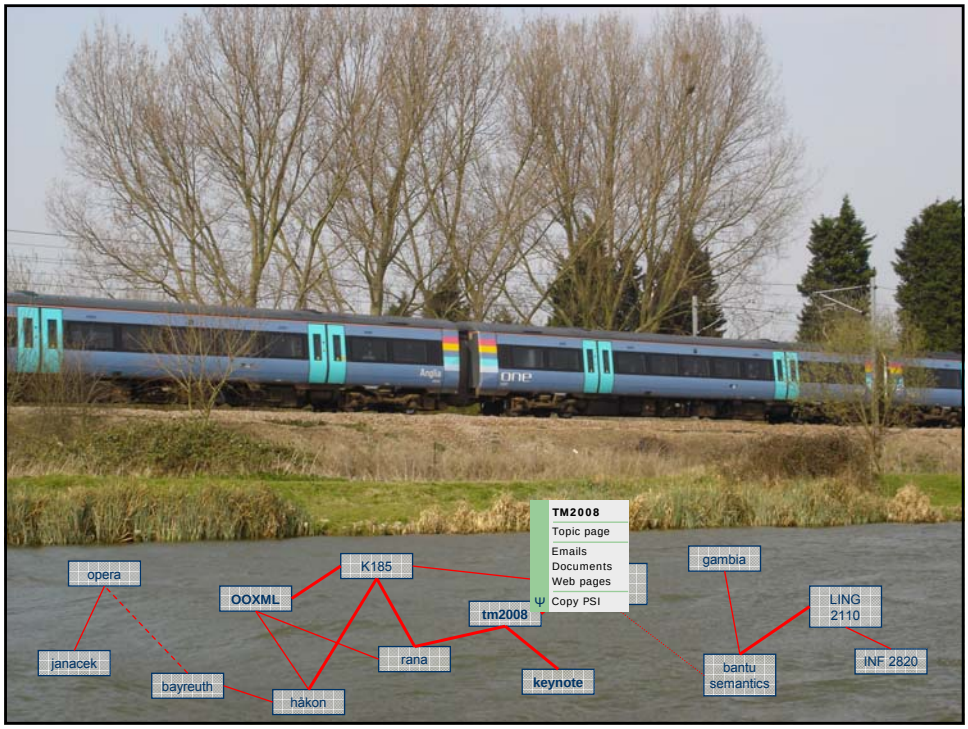
ONTOPEDIA
The Identity of Everything

Subject-centric desktop

- **I'm a Windows user**
- **Who uses Windows?**
  - Files in the file system
  - Outlook mail boxes
  - Browser bookmarks (favourites)
  - ...all thoroughly document centric...
- **Allow me to show you my desktop...**

http://www.topicmaps.com/tm2008/pepper.ppt


ONTOPEDIA


TM2008
Topic page
Emails
Documents
Web pages
Copy PSI
opera
K185
OOXML
tm2008
gambia
LING 2110
janacek
rana
keynote
bantu semantics
INF 2820
bayreuth
hakon

ONTOPEDIA
The Identity of Everything

Subject-centric file system

- **The file system is a hierarchy and that's a pain**
- **Trees aren't miscellaneous enough**
- **WinFS looked like it might change all that**
  - New data storage and management system announced in 2003
  - Didn't make it into Vista. Seems to have disappeared
- **Let the new file system be a topic map!**
  - "Folders" are topics with global identifiers
  - User-defined metadata on "folders" (internal occurrences)
  - External occurrences
  - Related through navigable, typed associations

`http://www.topicmaps.com/tm2008/pepper.ppt`

ONTOPEDIA
The Identity of Everything

Subject-centric operating system

- **Now that the file system is a topic map, why not go the whole hog?**
  - Services to applications for assigning PSIs
  - NLP based help for (semi-automatically) categorizing documents
  - Ability to extract fragments from the system topic map
  - Peer-to-peer features for exchanging fragments with others
  - Facilities for context-based virtual merges under user control
  - ...

`http://www.topicmaps.com/tm2008/pepper.ppt`

ONTOPEDIA
The Identity of Everything

The paradigm shift

- Topic Maps started out as a way to merge indexes
- It turned into a knowledge representation formalism
- But its significance is far greater

- Now the flag-bearer for subject-centric computing
- A paradigm shift in how we use computers

- Cf. object-oriented programming...
- ...and Copernicus

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Object-oriented programming

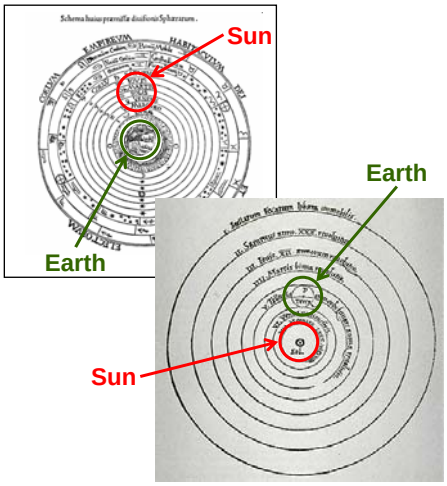
- Response to 1960's software crisis
  - Computer programs more and more complex
  - Difficult to maintain software quality
- Code simulates the world (as perceived by a human)
  - Objects represent real-world concepts (cf. topics)
  - They are grouped into classes (cf. topic types)
  - Data structures capture relationships between objects (cf. associations)
- Represented a paradigm shift in programming
  - OO languages now near universal (Java, C#, Ruby, Python, ...)

http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

The heliocentric revolution

- For 1,000s of years people thought that the sun revolved around the earth
- In 1543 Copernicus changed all that
- His heliocentric theory turned our understanding of the universe inside out.
- This was another paradigm shift

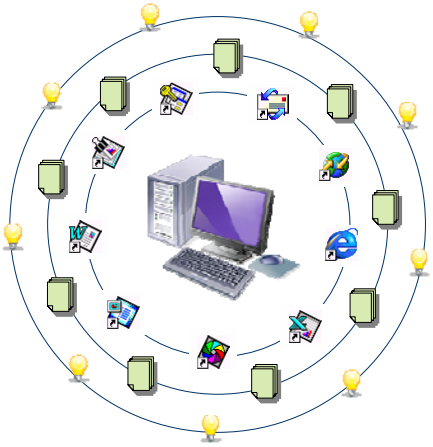


http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Subject-centric computing

- Today we face a similar situation in computing and information management
- Computers are at the centre of our information universe
- Applications and documents revolve around them
- The subjects we're really interested in are nowhere to be seen
- Or at least, nowhere to be found

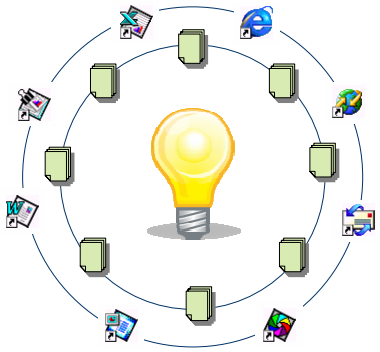


http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

Computing “as we may think”

- This is wrong, because it does not reflect how humans think
- Humans think in terms of subjects, concepts, ideas
- We must put subjects at the centre, because that’s what we’re really interested in
- This is the essence of subject-centric computing
- It really is a paradigm shift –
- Topic Maps is showing the way



http://www.topicmaps.com/tm2008/pepper.ppt

ONTOPEDIA
The Identity of Everything

THE END

Or is it the beginning?

- og forøvrig mener jeg at Norges nasjonale kunnskapsbase må baseres på emnekart...

http://www.topicmaps.com/tm2008/pepper.ppt