

Moving beyond 'Presentation vs. Content'

leveraging the wealth of semantic
categorization already at hand

The Semantic Web

... a “man-made woven web of data”
that **facilitates machines to**
understand the semantics, or
meaning, of information on the World
Wide Web.

The Semantic Web

Goal is to go “...**beyond linear presentation** of information (Web 1.0) and multi-linear presentation of information (Web 2.0) to make use of hyper-structures leading to entities of hypertext.”

http://en.wikipedia.org/wiki/Semantic_web

... “presentation of information” ...

“Presentation” entails **not only** when,
where, and how (i.e. physical structures)

but also that something *represents* something
else (i.e. conceptual structures)

Physical

Text
HTML
XML
CSS
WWW
Search Engines

Conceptual

Idea
Spider diagrams
XLST
?
Semantic Network(s)
?

Form

Text
HTML
XML
CSS
WWW
Search Engines

Function

Idea?
Spider diagrams?
XSLT
Thesaurus?
Semantic Network(s)
?

Re: Structures

Formal, **physical entities** build upward in hierarchical structures that allow for web crawling and searching

Conceptual structures are more complicated: **ontological** (i.e., this is a type of that) hierarchy **semiological** (i.e., this serves as a symbol for that)

In Practice

Physical structure example query:

Existence: *“Find all occurrences of X”*

Proximity: *“...where X occurs around Y”*

Conceptual structures example queries:

ontological: *“...where X belongs to Y”*

semiological: *“...where X implies Y”*

Implication = Nuance

A **'lossy'** system: **detail is reduced**

Detail is lost in **systematic** ways
(i.e. not completely haphazard)

Recapturing detail requires information that far
exceeds the amount that was lost

Implication = Nuance

Example: Translation

regnum cęlorum nulli pecunię potest
comparari: ne mag giuu(er)t(he)rid uuerthan:
tamen tanti ualet quantum habes. it mág tho
giuúnnian uuérthan. so mid mínneron so mid
méron. so man hauid

Overcoming challenges

Odd characters

regnum cęlorum nulli pecunię potest
comparari: ne mag giuu(er)t(he)rid uuerthan:
tamen tanti ualet quantum habes. it mág tho
giuúnnian uuérthan. so mid mínneron so mid
méron. so man hauid

Overcoming challenges

Unicode or replacement rules

regnum caelorum nulli pecuniae potest
comparari: ne mag giuu(er)t(he)rid uuerthan:
tamen tanti ualet quantum habes. it mág tho
giuúnnian uuérthan. so mid mínneron so mid
méron. so man hauid

Overcoming challenges

Mark-up: linkability, searchability, etc.

regnum celorum nulli pecunie potest
comparari: ne mag giuu(er)t(he)rid uuerthan:
tamen tanti ualet quantum habes. it mág tho
giuúnnian uuérthan. so mid mínneron so mid
méron. so man hauid

Overcoming challenges

Mark-up 2: call out special features

[regnum cęlorum nulli pecunię potest
comparari:]_{LAT} [ne mag giuu(er)t(he)rid
uuerthan:]_{OS} [tamen tanti ualet quantum
habes.]_{LAT} [it mág tho giúúnnian uuérthan. so mid
mínneron so mid méron. so man hauid]_{OS}

Overcoming challenges

Mark-up 3: syntactic hierarchy



nulli pecunię potest comparari

Overcoming challenges

Mark-up 4: paradigmatic semantic similarity

nulli pecunię potest comparari
divitiae
rei
bonae

Ongoing Challenges

NEED: syntagmatic semantic similarity

(i.e. syntax influences meaning, but not always in the same ways within or between languages)

tamen tanti ualet quantum habes.

m.sg.gen



m.sg.acc

2.sg.pres.ind

so mid mínneron so mid méron. so man hauid

prep. m.pl.dat.

prep. m.pl.dat.

3.sg.pres.ind

Ongoing Challenges

NEED: Connotation, euphamism, metaphor

nec ante nec post eam **cognoscebat.**
ne **uuas mit iru.**

ita ut **cum ea**
ne **ualctimo sia**

Ongoing Challenges

NEED: Connotation, euphemism, metaphor

nec ante nec post eam cognoscebat.
ne uuas mit iru.

ita ut cum ea
ne ualctimo sia

Ongoing Challenges

NEED: Connotation, euphamism, metaphor

nec ante nec post eam **cognoscebat.**
ne **uuas** mit iru.

ita ut **cum** ea
ne ualcta imo sia

Continuing Hurdles

Digitize: stand. etym. dictionaries (OCR)

Varying standard/non-standard abbrev.

Varying citation schemes

Multi-lingual, non-narrative texts

***walg?**, ahd., (Adj.): Vw.: s. ana-; E.: germ. *walga- (2), *walgaz, Adj., sich drehend?; s. idg. *uolg-, V., drehen, sich bewegen, Pokorny 1144; vgl. idg. *uel- (7), *uelə-, *ulē-, V., drehen, winden, wälzen, Pokorny 1140

***walgī?**, ahd., (st. F.) (i): Vw.: s. ana-; E.: germ. *walgī-, *walgīn, sw. F. (n), Fülle; s. idg. *uolg-, V., drehen, sich bewegen, Pokorny 1144; vgl. idg. *uel- (7), *uelə-, *ulē-, V., drehen, winden, wälzen, Pokorny 1140

walgōn* 5, walagōn*, ahd., sw. V. (2): nhd. wälzeln?, wälzen, sich wälzen, rollen, wirbeln, bewegen; ne. walk (V.)?, roll (V.); ÜG.: lat. (ambulare) Gl, volvere Gl; Vw.: s. bi-; Q.: Gl (10. Jh.); E.: germ. *walg-, V., wälzen; idg. *uolk-, V., drehen, bewegen, Pokorny 1144; s. idg. *uel- (7), *uelə-, *ulē-, V., drehen, winden, wälzen, Pokorny 1140; W.: s. mhd. walgen, sw. V., rollen (intr.), bewegen, wimmeln, wälzen (tr.), rollen; nhd. (ält.) walgen, sw. V., wälzen, rollen, sich wälzen, DW 27, 1227

walka* 2, walca*, ahd., sw. F. (n)?; nhd. Band (N.), Fallstrick, Schlinge, Fessel (F.) (1); ne. rope (N.), snare (N.); ÜG.: lat. pedica Gl, vitta Gl; Q.: Gl (12. Jh.); I.: lat. beeinflusst?; E.: s. germ. *walkan, st. V., drehen, wandern, walken; idg. *uolg-, V., drehen, sich bewegen, Pokorny 1144; vgl. idg. *uel- (7), *uelə-, *ulē-, V., drehen, winden, wälzen, Pokorny 1140

walkan* (1) 1, walcan*, ahd., red. V.: nhd. walken, verfilzen, verdichtet (= giwalkan), verfilzt (= giwalkan); ne. full (V.), felt (V.), felted (= giwalkan); ÜG.: lat. concretus (= giwalkan) Gl; Vw.: s. fir-; Hw.: s. lang. *walkan; Q.: Gl (11. Jh.), ON; E.: germ. *walkan, st. V., drehen, wandern, walken; idg. *uolg-, V., drehen, sich bewegen, Pokorny 1144?; s. idg. *uel- (7), *uelə-, *ulē-, V., drehen, winden, wälzen, Pokorny 1140?; W.: s. nhd. walken, sw. V., hin und her bewegen, in der Walke bearbeiten, DW 27, 1244

***walkan?**, lang., st. V.: nhd. walken, pressen; ne. full (V.), press (V.); Hw.: s. ahd. *walkan; Q.: it. gualcare, walken

***walkan (2)?**, *walcan, ahd.? (Part. Prät.=Adj.): Vw.: s. fir-, gi-; Hw.: s. walkan* (1) (red. V.)

walkāri 7, walcāri*, ahd., st. M. (ja): nhd. Walker, Tuchbereiter; ne. fuller; ÜG.: lat. coagitator Gl, compressor Gl, fullo Gl; Q.: Gl (12. Jh.); I.: Lsch. lat. fullo?; E.: s. germ. *walkan, st. V., drehen, wandern, walken; idg. *uolg-, V., drehen, sich bewegen, Pokorny 1144?; vgl. idg. *uel- (7), *uelə-, *ulē-, V., drehen, winden, wälzen, Pokorny 1140; W.: mhd. walker, st. M., Walker; nhd. Walker, M., Walker, der Tuche und Felle walkt, DW 27, 1250

***walkjan**, lang., sw. V.?; nhd. walken, pressen; ne. full (V.); Hw.: s. ahd. walkan* (1); Q.: moden. sgualzir, walken, sien. gualcire, walken, it. gualcire, übel zurichten, zerknittern, amasen. gualcire, den Flachs rotten, mailänd. magolciá, abtasten, irpin. sguanciá, abgreifen, zerknüllen

walko* 1, walco*, ahd.?, sw. M. (n): nhd. Stampfer?; ne. pounder?; ÜG.: lat. (pinzonaticus) Gl; Vw.: s. fahs-; Q.: Gl (13. Jh.); I.: lat. beeinflusst?; E.: s. walkan (1)

walkwurf* 1, walcwurf*, ahd., st. M. (i): nhd. »Walkwurf«, »Haarziehen«; ne. pulling (N.); ÜG.: lat. (discriminalia eicere de capite) LBai; Q.: LBai (vor 743); I.: lat. beeinflusst?; E.: s. wurf; s. germ. *walkan, st. V., drehen, wandern, walken; idg. *uolg-, V., drehen, sich bewegen, Pokorny 1144; vgl. idg. *uel- (7), *uelə-, *ulē-, V., drehen, winden, wälzen, Pokorny 1140

walc..., ahd.: Vw.: s. walk...

walcatorium* 2 und häufiger, lat.-ahd.?, N.: nhd. Walkmühle; ne. fulling-mill; Q.: Urk (875); E.: s. germ. *walkan, st. V., drehen, wandern, walken; idg. *uolg-, V., drehen, sich bewegen, Pokorny 1144; vgl. idg. *uel- (7), *uelə-, *ulē-, V., drehen, winden, wälzen, Pokorny 1140

wald 50?, ahd., st. M. (i): nhd. Wald, Wildnis, Baumwipfel?, Wüste, Einöde; ne. woodland; ÜG.: lat. desertum (= wuostinna waldes) O, eremus B, Gl, (Libanus) NGl, nemus Gl, N. saltus (M.) (2) Gl, silva Gl, N, NGl, Ph; Vw.: s. sīt-; Hw.: s. lat.-ahd.? gualdus, lang. *wald; vgl. anfrk. wald* (1), as. wald* (1); Q.: B, GB, Gl (765), LN, N, NGl, O, Ph, ON, PN, WS; E.: germ. *walþu-, *walþuz, st. M. (u), Wald, Wildnis, Heide (F.) (1); s. idg. *uel- (4), *uelə-, Sb., Haar (N.), Wolle, Gras, Ähre, Wald, Pokorny 1139; W.: mhd. walt, st. M., Wald, Waldgebirge, Baumbestand, Waldhöhle; nhd. Wald, M., Wald, DW 27, 1072; Son.: Sachglr = Sachglossar Hermeneumata (Vocabularius St. Galli) (Sankt Gallen, Stiftsbibliothek 913), Tgl09 = Glossen zum Jeremias-Kommentar des Hieronymus (München, Bayerische Staatsbibliothek Clm 14425)

Goals

Create a sort of meta-crawler that:

- Is capable of reading a text 'as is'
(in line with [Berners-Lee's 1980/1989](#) requirement for "access to existing data")
- Builds network between the texts
 - Link based on similarity of meaning
 - Reduce results based on sim. of form
 - Create XML semantic, etymol. info

Benefits

- Mark up w/ potentially infinite semantic info for each item
- Computer-driven, not reliant on programmer's or user's knowledge
- Quickly grow semantically marked-up materials, essentially taking the process of making "man-made woven web of data" from a **sewing machine** to a semantic **loom**

Benefits

- Potentially create physical structure (XML, syntax trees, etc.) automatically
- Begin unlocking pragmatics (nuance, metaphor, etc.)
to computation

Teilprojekt: “Sprachliche Beziehungen” Sprachdatenbank ‘**Simplex**’