

Towards the Semantic Web

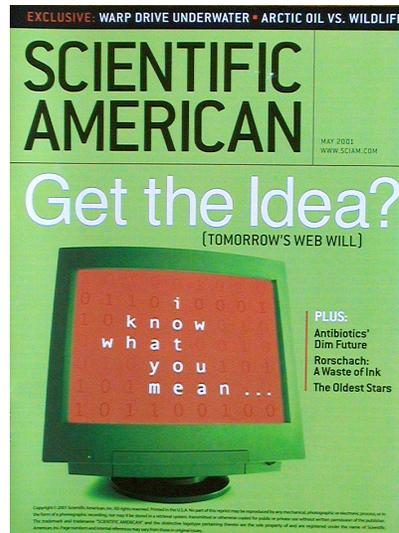
Ora Lassila

Research Fellow, Nokia Research Center (Boston)
Chief Scientist, Nokia Venture Partners LLP
Advisory Board Member, W3C

W3C Semantic Tour, London, June 2003

Towards the Semantic Web

- **Motivation**
- **Technologies**
 - representation formats
 - ontologies
 - Web Services
- **Future**
- **Q & A**



Motivation for the Semantic Web

- **Original driver: Automation**
 - it would be nice if computers could do more (on the Web)
 - solution: make information on the Web more “machine-friendly”
 - origins of the Semantic Web are in web metadata
- **Short term goal: Interoperability**
 - combining information from multiple sources
 - Web Services: discovery, composition
 - “serendipitous” interoperability
- **Long term goal: “Departure from the Tool Paradigm”**
 - instead of using computers like tools, make them work *on our behalf*
 - removing humans from the loop to the extent possible

More About the Motivation

- **Problem: Web was built for humans**
 - human interpretation needed to “understand” content (this does not scale)
 - consequently, automation is difficult
 - it is particularly difficult to automate “unforeseen” situations
- **Rough solution: make the Web friendlier for machines**
 - we need “machine-understandable” content (not “machine-readable”, we already have that)
 - “machine-understandable” means content with *accessible formal semantics*
- **The Web is more than just a “library”**
 - think of it as infrastructure for services & functionality
- **Drivers**
 - automation (e.g., in search), interoperability (e.g., in e-commerce)
 - but: compelling business models may still be missing

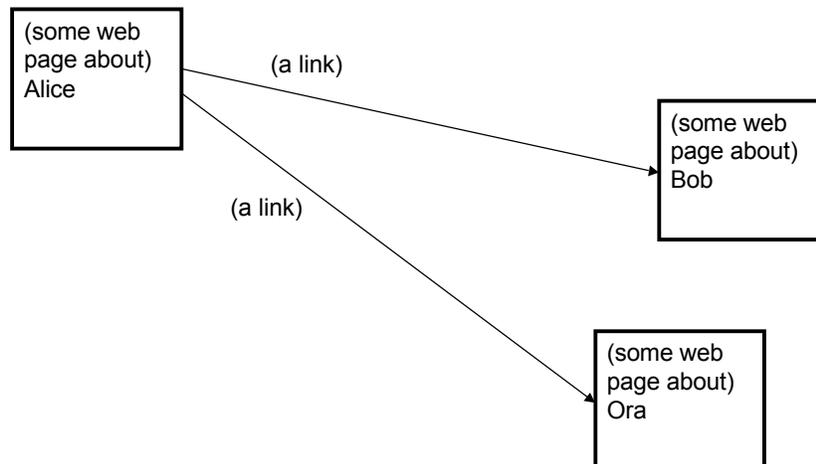
WWW: an Architecture for Linkages

- **Current Web architecture essentially gives us a framework for “pointing”**
- **Problem is that this pointing has no meaning**
 - (except sometimes through human interpretation)

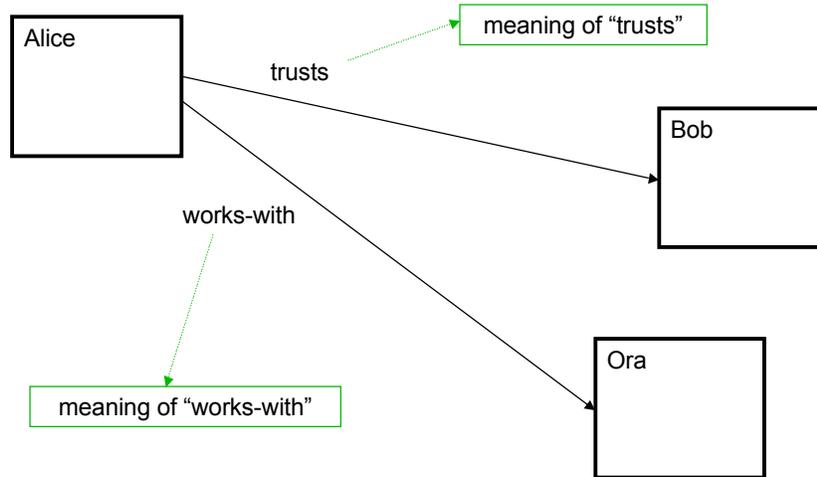
Can we improve on this?

Note: for us (humans), separating our own interpretation from (largely syntactic) representation is hard

Linkages on the “Old Web”



Linkages on the “Semantic Web”



Linkages on the “Semantic Web” (2)

- **Semantic Web resources (the “nodes”)** can
 - stand alone, or
 - denote other things (e.g., physical entities)
- **Hypertexts become “semantic” networks**
 - this is good for agents and automation
 - e.g., semantic navigation of hypertexts
 - how does one “name” the semantic links and nodes?

Semantics via Sharing

- **Controlled vocabularies**
 - interoperability improves if the same term is always used to denote the same thing (e.g., instead of arbitrary keywords, choose from a list)
- **What is an “ontology”**
 1. a controlled vocabulary
 2. a concept taxonomy
 3. other relations between concepts
 - definition: “A specification of conceptualization” (Gruber)
- **Library scientists are good with this stuff**
 - e.g., Dewey Decimal System is an ontology

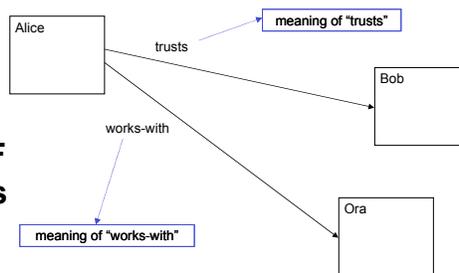
More about Ontologies

- **How to build ontologies?**
 - we could form committees...
 - the Dublin Core initiative took years to decide on 15 core metadata elements
 - (this doesn't mean that DC wouldn't be a good standard)
 - my preference is the “Darwinian” approach
 - good and/or popular ontologies will prevail
 - we must have a framework which allows ontology extension (RDF does)
 - probably some combination of official standards and de-facto standards is the way to go
- **Some “upper ontology” projects underway**
- **Ontologies enable reasoning**
 - this allows the move from “syntactic” to “semantic” processing
 - but: where does “semantic data” come from?

Resource Description Framework



- **Originally conceived as W3C's metadata model**
 - document metadata for digital libraries, content rating, site maps, etc.
 - normative reference: Lassila & Swick, "Resource Description Framework Model and Syntax Specification", W3C Recommendation, 1999
- **RDF has**
 - a data model of directed labeled graphs (DLGs)
 - an XML-based syntax for serializing DLGs
- **Nodes & arcs in an RDF DLG are named by URIs**
 - important for robust vocabulary creation



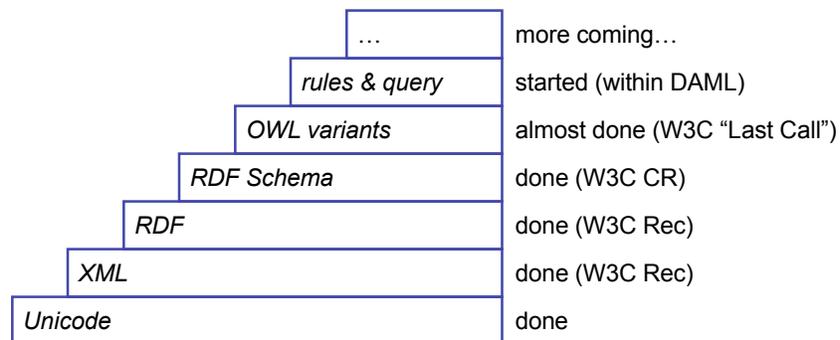
“It’s a Model, Stupid!”

- **Simple data model**
 - think of it either as *directed labeled graphs* or in object-oriented terms
 - more powerful than the trees XML gives you
- **Graphs decompose into object/attribute/value -triples**
 - “subject/predicate/object” = a statement
 - (in RDF parlance, nodes are called “resources” and arcs “properties”)
- **Everything in an RDF graph is named by URIs**
 - when naming is not based on mere words, name conflicts can be avoided
 - graphs can span multiple hosts (servers, etc.)
- **RDF is followed by more powerful languages**
 - DAML+OIL (from the DARPA Agent Markup Language program)
 - OWL (from W3C's WebOnt working group)

W3C Web Ontology Language (OWL)

- Layered on RDF, offers more expressive power
- Comes in three “flavors” (layered on one another)
 - OWL Full
 - OWL DL
 - OWL Lite
- **OWL DL is particularly attractive for “real world” applications**
 - complete & decidable
 - good reasoning engines exist (e.g., from Network Inference)
- **OWL Lite is what RDF Schema should have been**
 - simple yet expressive language for concept taxonomies
- **Standardization work almost done**

Stepping Towards the Semantic Web



- **Semantic Web is built in a layered manner**
- **Not everybody needs all the layers**

Isn't It Enough to Just Use XML?

- **Short answer: no**
 - the typical answer – albeit incorrect – is “yes”
- **Long answer: XML offers a way to introduce new syntax (new names, tags, ...), but no way of introducing or coordinating semantics**
- **XML has a tree-like data model**
 - if your (representational) problem does not lend itself to be a tree, you lose (sorry – this is even before we get to the “semantics” part)
- **Hype (from a major IT company white paper): “The industry is clearly focusing in on [XML] as the *lingua franca* to enable Web services...”**
 - not only is XML not a *lingua franca*, it is not even a *lingua*

XML ≠ Machine Accessible Meaning (1)

林克昌 根留台灣 可能增高

在愛戴者熱心奔走之下，華裔指揮家林克昌根留台灣的可行性又提升了幾分。兩廳院主任李炎、國家音樂廳樂團副團長黃奕明日前親赴林克昌、石聖芳寓所拜會，並提出多場客座邀約。此外，台灣省立交響樂團團長陳澄維也早「下訂」，邀請林克昌赴台中演奏，從八月十日起訓練省交，為期長達一個月。

在台灣諸多公家樂團中，陳澄維是以實際行動表達對林克昌肯定的樂界人士之一，曾多次公開表示對林克昌指揮才華的讚賞，而且幾乎每個樂季都邀請林克昌客座演出。

此外，林克昌上個月赴俄羅斯與頂尖的「俄羅斯國家管絃樂團」灌錄了柴可夫斯基晚期三大交響曲以及「羅密歐與茱麗葉」、「斯拉夫進行曲」、「義大利隨想曲」，最後的DAT母帶也在前兩天寄回台灣。製作人楊忠衛與林克昌試聽之後，對錄音效果—尤其音質表現感到相當滿意，楊忠衛估計呈現了七分林克昌指揮神韻。

俄羅斯國家管絃樂團首席布魯尼日前也讚譽林克昌的指揮藝術有三大特點：一是控制自如的彈性速度；二是強烈的動感對比；三是宛如呼吸歌場的旋律處理。這些對錄音師而言都構成很大挑戰。俄國錄音師雖然採用多軌混音，但定位、場面都有可觀之處。

- (thanks to Frank van Harmelen, VUA)

XML ≠ Machine Accessible Meaning (2)

林克昌 根留台灣 可能增高

在愛戴者熱心奔走之下，華裔名指揮家林克昌根留台灣的可行性又提升了幾分。兩廳院主任李炎、國家音樂廳樂團副團長黃奕明日前親赴林克昌、石聖芳寓所拜會，並提出多場客座邀約。此外，台灣省立交響樂團團長陳澄雄也早早「下訂」，邀請林克昌赴台中聽峰，從八月十日起訓練省交，為期長達一個月。

在台灣諸多公家樂團中，陳澄雄是以實際行動表達對林克昌肯定的樂界人士之一，曾多次公開表示對林克昌指揮才華的欽佩，而且幾乎每個樂季都邀請林克昌客座演出。

此外，林克昌上個月赴俄羅斯與頂尖的「俄羅斯國家管絃樂團」灌錄了柴可夫斯基晚期三大交響曲以及「羅密歐與茱麗葉」、「斯拉夫進行曲」、「義大利隨想曲」，最後的DAT母帶也在前兩天寄回台灣。製作人楊忠衛與林克昌試聽之後，都對錄音效果—尤其音質表現感到相當滿意，楊忠衛估計呈現了七分林克昌指揮神韻。

俄羅斯國家管絃樂團首席布魯尼日前也讚譽林克昌的指揮藝術有三大特點：一是控制自如的彈性速度；二是強烈的動感對比；三是宛如呼吸歌嗎的旋律處理。這些對錄音師而言都構成很大挑戰。俄國錄音師雖然採用多軌混音，但定位、場面都有可觀之處。

name

education

work

personal

CV

XML ≠ Machine Accessible Meaning (3)

林克昌 根留台灣 可能增高

在愛戴者熱心奔走之下，華裔名指揮家林克昌根留台灣的可行性又提升了幾分。兩廳院主任李炎、國家音樂廳樂團副團長黃奕明日前親赴林克昌、石聖芳寓所拜會，並提出多場客座邀約。此外，台灣省立交響樂團團長陳澄雄也早早「下訂」，邀請林克昌赴台中聽峰，從八月十日起訓練省交，為期長達一個月。

在台灣諸多公家樂團中，陳澄雄是以實際行動表達對林克昌肯定的樂界人士之一，曾多次公開表示對林克昌指揮才華的欽佩，而且幾乎每個樂季都邀請林克昌客座演出。

此外，林克昌上個月赴俄羅斯與頂尖的「俄羅斯國家管絃樂團」灌錄了柴可夫斯基晚期三大交響曲以及「羅密歐與茱麗葉」、「斯拉夫進行曲」、「義大利隨想曲」，最後的DAT母帶也在前兩天寄回台灣。製作人楊忠衛與林克昌試聽之後，都對錄音效果—尤其音質表現感到相當滿意，楊忠衛估計呈現了七分林克昌指揮神韻。

俄羅斯國家管絃樂團首席布魯尼日前也讚譽林克昌的指揮藝術有三大特點：一是控制自如的彈性速度；二是強烈的動感對比；三是宛如呼吸歌嗎的旋律處理。這些對錄音師而言都構成很大挑戰。俄國錄音師雖然採用多軌混音，但定位、場面都有可觀之處。

<name>

<education>

<work>

<personal>

<CV>

XML ≠ Machine Accessible Meaning (4)

林克昌 根留台灣 可能增高

在愛戴者熱心奔走之下，華裔名指揮家林克昌根留台灣的可行性又提升了幾分。兩廳院主任李炎、國家音樂廳樂團副團長黃奕明日前親赴林克昌、石聖芳寓所拜會，並提出多場客席邀約。此外，台灣省立交響樂團團長陳澄雄也早早「下訂」，邀請林克昌赴台中聽峰，從八月十日起訓練省交，為期長達一個月。

在台灣諸多公家樂團中，陳澄雄是以實際行動表達對林克昌肯定的樂界人士之一，曾多次公開表示對林克昌指揮才華的欽佩，而且幾乎每個樂季都邀請林克昌客席演出。

此外，林克昌上個月赴俄羅斯頂尖的「俄羅斯國家管絃樂團」灌錄了柴可夫斯基晚期三大交響曲以及「羅密歐與茱麗葉」、「斯拉夫進行曲」、「義大利進行曲」，最後的DAT母帶也在前兩天寄回台灣。製作人楊忠衛與林克昌試聽之後，都對錄音效果—尤其音質表現感到相當滿意，楊忠衛估計呈現了七分林克昌指揮神韻。

俄羅斯國家管絃樂團首席布魯尼日前也讚譽林克昌的指揮藝術有三大特點：一是控制自如的彈性速度；二是強烈的動感對比；三是宛如呼吸歌鳴的旋律處理。這些對錄音師而言都構成很大挑戰。俄國錄音師雖然採用多軌混音，但定位、場面都有可觀之處。

The diagram illustrates the mapping of XML-like structures to the text. It shows several instances of <>[]> symbols. Red arrows point from these symbols to specific lines of text: one points to the title, another to the first paragraph, a third to the second paragraph, a fourth to the third paragraph, and a fifth to the fourth paragraph. A large red bracket on the right side groups these symbols together, with a final <>[]> symbol positioned below the bracket.

Reasoning and Inference

- Reasoning allows one to draw inferences based on generalized “rules”
 - generation of “more” semantic information
 - simplest practical form: polymorphism in OO systems
- Enabled by ontologies
- Reasoning eases interoperability
 - relationships between different but compatible ontologies & data could be inferred

Reasoning example:

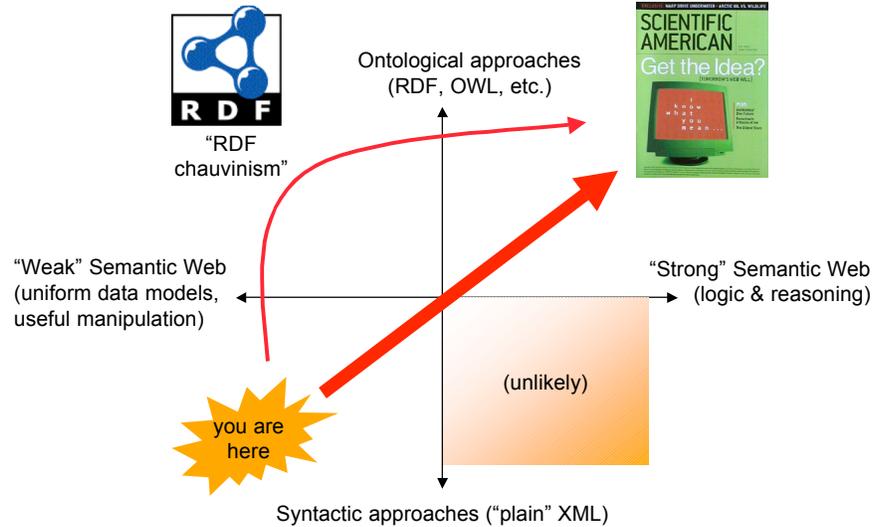
1. X is a Cat
2. a Cat is a Mammal
3. a Mammal gives birth to live young

□

- X gives birth to live young

Note: This is AI

Semantic Web: Characterizations



21 ©NOKIA 7/14/03 - Ora Lassila

NOKIA

Interoperability of Services

- **Semantic Web, via ontologies and reasoning, will improve interoperability of information systems**
- **This can be applied to “services”**
 - semantic description of service interfaces enables automatic discovery, composition, etc.
 - DARPA’s DAML-S activity (Stanford, CMU, Yale, SRI, BBN, Nokia)
 - analog to “Tower of Babel” (from Genesis 11:1-9)
 - will Web Services succeed without the Semantic Web? (I think not)
- **Substitution of “equivalent” services**
- **Web Services are a good abstraction of all kinds of functionality**

22 ©NOKIA 7/14/03 - Ora Lassila

NOKIA

Tools & Beyond (examples)

Tools

- hammer & nails
- calendaring software
- e.g., Google
- almost any software today

Beyond tools

- building contractor
- automated “secretary”
- answers from a “search agent”
- various personal assistants...

Fulfillment of the Vision (Using AI)

- **Knowledge representation**
 - (obvious: the Semantic Web is all about KR)
 - formal semantics as “the Manifest Destiny of AI”
- **Automated planning**
 - enables autonomous operation
 - useful in many tasks (e.g., service composition)
- **Machine learning**
 - enables adaptivity
 - could be used in bootstrapping semantic annotations for existing content
- **The “AI Paradox”**
 - well-understood things stop being AI (e.g., OOP, rules, logic)
 - parallels between AI and the Semantic Web: the latter also has aspects which, once adopted, will stop being “Semantic Web”

Summary

- **Use of human interpretation does not scale**
- **We need to**
 - move from tools to autonomous systems that work on our behalf
 - introduce formal semantics (machine-understandable content)
- **Ontologies □ Reasoning □ Agents**
 - we have only done the first step and started on the second...
 - (business models for all this are needed)
- **We may need artificial intelligence to ultimately fulfill the Semantic Web vision**
 - (some of you may have been misinformed about this earlier)

Questions?

- <http://www.nokia.com/research/semanticweb>
- <mailto:ora.lassila@nokia.com>