Knowledge Representation in Practice: Project Halo and the Semantic Web

Mark Greaves
Vulcan, Inc.
markg@vulcan.com (206) 342-2276

Talk Outline

- **The Halo Vision**
- **Systems AI – Vulcan's Halo Program**
  - The Halo Pilot: The Limits of Expert Systems
  - Halo Phase II: Deep Reasoning over the AP problem
  - Halo Today: Leveraging the Web
- **The Future of Halo**

Talk Outline

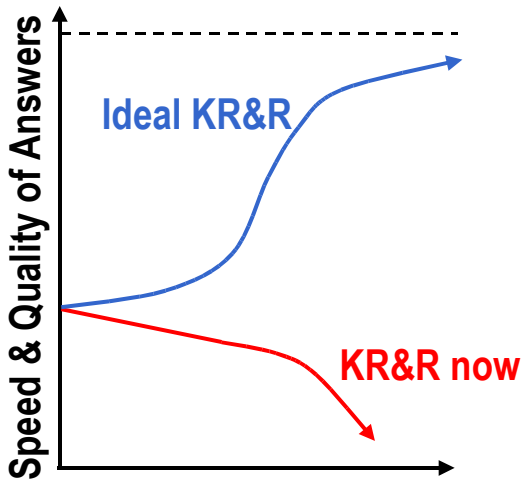
- **The Halo Vision**
- **Systems AI – Vulcan's Halo Program**
  - The Halo Pilot: The Limits of Expert Systems
  - Halo Phase II: Deep Reasoning over the AP problem
  - Halo Today: Leveraging the Web
- **The Future of Halo**

KR&R Systems, Scaling, and the Google Property

- **We seek KR&R systems that have the "Google Property:"** ***they get (much) better as they get bigger***
  - Google's PageRank™ yields better relevance judgments when it indexes more pages
  - Current KR&R systems have the antithesis of this property

- **So what are the components of a scalable KR&R system?**
  - Distributed, robust, reliable infrastructure
  - Multiple linked ontologies and points of view
    - Single ontologies are feasible only at the program/agency level
  - Mixture of deep and shallow knowledge repositories
  - Simulations and procedural knowledge components
    - "Knowing how" and "knowing that"
  - Embrace uncertainty, defaults, and nonmonotonicity in all components
  - Uncertainty in the KB – you don't know what you know, things go away, contradiction is rampant, resource-aware computing is necessary, surveying the KB is not possible

KR&R Goals



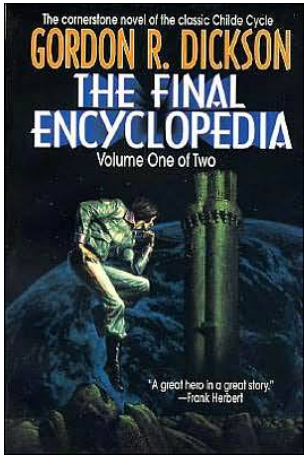
Scalable KR&R Systems should look just like the Web!!

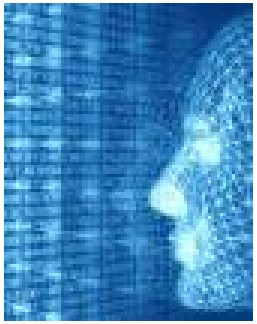
(coupled with great question-answering technology)

VULCAN

Envisioning the Digital Aristotle for Scientific Knowledge

- **Inspired by Dickson's Final Encyclopedia, the HAL-9000, and the broad SF vision of computing**
  - The "Big AI" Vision of computers that work with people
- **The volume of scientific knowledge has outpaced our ability to manage it**
  - This volume is too great for researchers in a given domain to keep abreast of all the developments
  - Research results may have cross-domain implications that are not apparent due to terminology and knowledge volume
- **"Shallow" information retrieval and keyword indexing systems are not well suited to scientific knowledge management because they cannot reason about the subject matter**
  - Example: "What are the reaction products if metallic copper is heated strongly with concentrated sulfuric acid?" (Answer: Cu^{2+}, $SO_2(g)$, and H_2O)
- **Response to a query should *supply the answer* (possibly coupled with conceptual navigation) rather than simply list 1000s of possibly relevant documents**





VULCAN

How do we get to the Digital Aristotle?

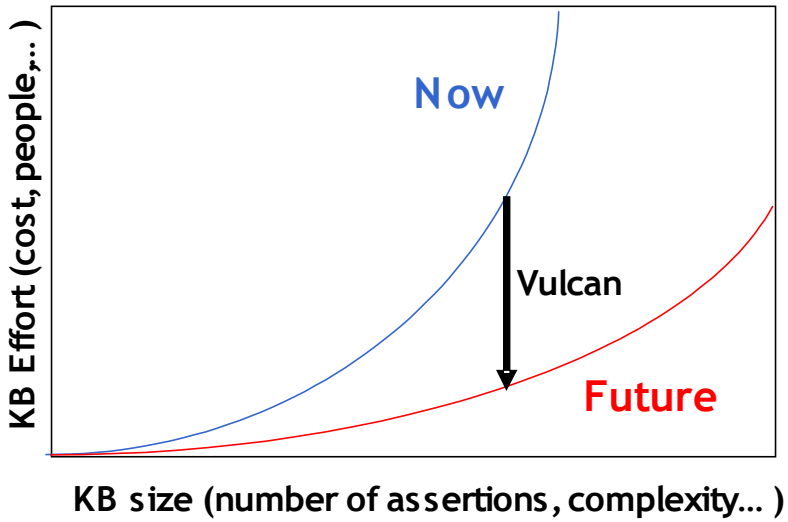
- **What we want:**
  - Technology to enable a global, widely-authored, very large knowledge base (VLKB) about human affairs and science,
  - Technology that answers questions and proactively supplies information,
  - Technology that uses powerful reasoning about rules and processes, and
  - Technology that can be customized in its content and actions for individual organizations or people

How do we get to the Digital Aristotle?

- **What we want:**
  - Technology to enable a global, widely-authored, very large knowledge base (VLKB) about human affairs and science,
  - Technology that answers questions and proactively supplies information,
  - Technology that uses powerful reasoning about rules and processes, and
  - Technology that can be customized in its content and actions for individual organizations or people

- **Vulcan's Goals**
  - Address the problem of **scale** in Knowledge Bases
    - Scaling by web-style participation
    - Incorporate large numbers of people in KB construction and maintenance
  - Have high **impact**
    - Show that the Digital Aristotle is possible
    - Change our experience of the Web
    - Have quantifiable, explainable metrics
  - Be a **commercializable** approach

- **Project Halo is a concrete research program that addresses these goals**



PROJECT HALO®

VULCAN

Talk Outline

- **The Halo Vision**
- **Systems AI – Vulcan's Halo Program**
  - The Halo Pilot: The Limits of Expert Systems
  - Halo Phase II: Deep Reasoning over the AP problem
  - Halo Today: Leveraging the Web
- **The Future of Halo**

The Project Halo Pilot (2004)

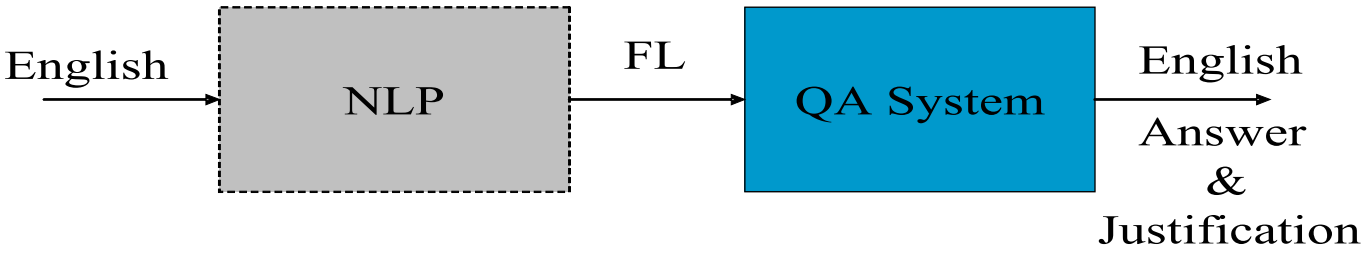
- **In 2004, Vulcan funded a six-month effort to determine the state-of-the-art in fielded "deep reasoning" systems**
  - Can these systems support reasoning in scientific domains?
  - Can they answer novel questions?
  - Can they produce domain appropriate answer justifications?
- **Three teams were selected, and used their available technology**
  - SRI, with Boeing Phantom Works and UT-Austin
  - Cycorp
  - Ontoprise GmbH

SRI International

ontoprise
know how to use Know-how

cycorp

- **No NLP in the Pilot**



VULCAN

The Halo Pilot Domain

- **70 pages from the AP-chemistry syllabus (Stoichiometry, Reactions in aqueous solutions, Acid-Base equilibria)**
  - Small and self contained enough to be do-able in a short period of time, but large enough to create many novel questions
  - Complex “deep” combinations of rules
  - Standardize exam with well understood scores (AP1-AP5)
  - Chemistry is an exact science, more “monotonic”
  - No undo reliance on graphics (e.g., free-body diagrams)
  - Availability of experts for exam generation and grading

- **Example: Balance the following reactions, and indicate whether they are examples of combustion, decomposition, or combination**
  - $C_4H_{10} + O_2 \rightarrow CO_2 + H_2O$
  - $KClO_3 \rightarrow KCl + O_2$
  - $CH_3CH_2OH + O_2 \rightarrow CO_2 + H_2O$
  - $P_4 + O_2 \rightarrow P_2O_5$
  - $N_2O_5 + H_2O \rightarrow HNO_3$

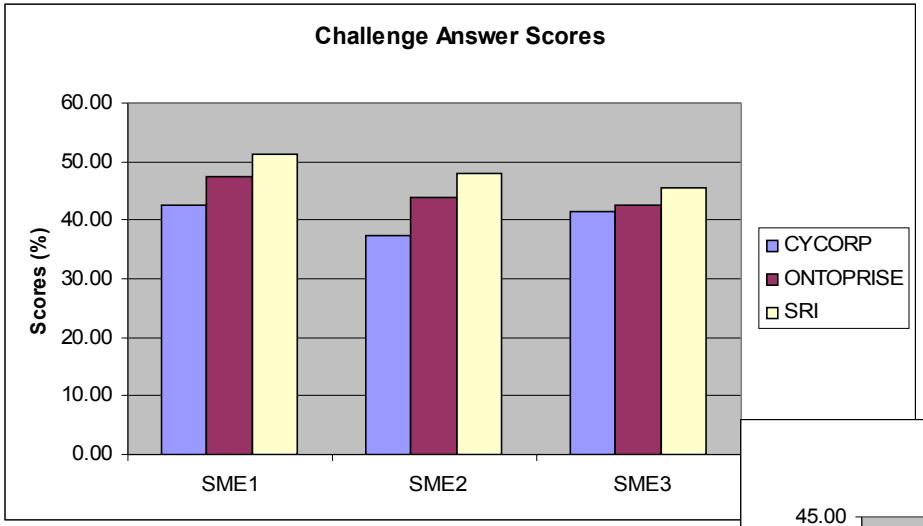
VULCAN

Halo Pilot Evaluation Process

- **Evaluation**
  - Teams were given 4 months to formulate the knowledge in 70 pages from the AP Chemistry syllabus
  - Systems were sequestered and run by Vulcan against 100 novel AP-style questions (hand coded queries)
  - Exams were graded by chemistry professors using AP methodology

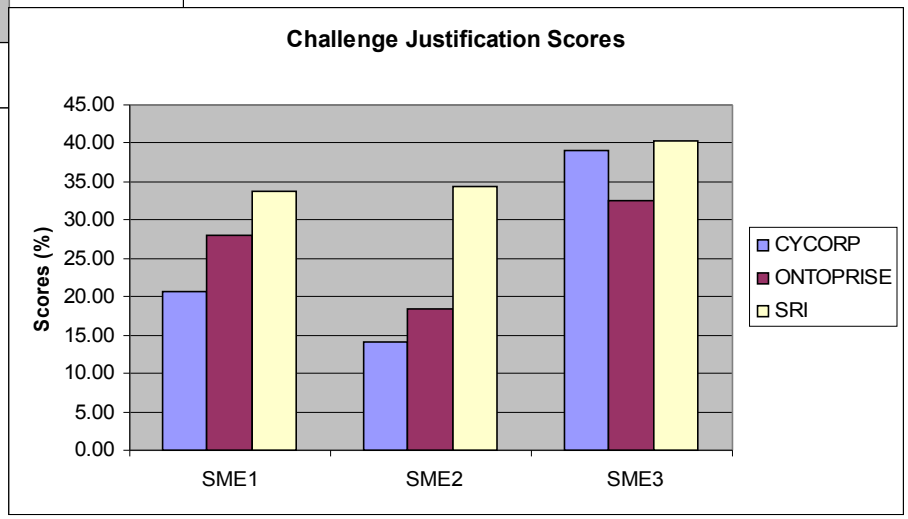
- **Metrics**
  - **Coverage:** The ability of the system to answer novel questions from the syllabus
    - What percentage of the questions was the system capable of answering?
  - **Justification:** The ability to provide concise, domain appropriate explanations
    - What percentage of the answer justifications were acceptable to domain evaluators?
  - **Query encoding:** The ability to faithfully represent queries
  - **Brittleness:** What were the major causes of failure? How can these be remedied?

Halo Pilot Results



Best scoring system achieved roughly an AP3 (on our very restricted syllabus)

Cyc had issues with answer justification and question focus



Full Details in *AI Magazine* 25:4, "Project Halo: Towards a Digital Aristotle"
...and at www.projecthalo.com

VULCAN

Talk Outline

- **The Halo Vision**
- **Systems AI – Vulcan's Halo Program**
  - The Halo Pilot: The Limits of Expert Systems
  - Halo Phase II: Deep Reasoning over the AP problem
  - Halo Today: Leveraging the Web
- **The Future of Halo**

VULCAN

From the Halo Pilot to the Halo Project

- **Halo Pilot Results**
  - Much better than expected results on a very tough evaluation
  - Most failures attributed to modeling errors due to contractors' lack of domain knowledge
  - Expensive: O($10,000) per page, per team

- **Project Halo Goal: To determine whether tools can be built to facilitate robust knowledge formulation, query and evaluation by domain experts, with ever-decreasing reliance on knowledge engineers**
  - Can SMEs build robust question-answering systems that demonstrate excellent coverage of a given syllabus, the ability to answer novel questions, and produce readable domain appropriate justifications using reasonable computational resources?
  - Will SMEs be capable of posing questions and complex problems to these systems?
  - Do these systems address key failure, scalability and cost issues encountered in the Pilot?

- **Scope: Selected portions of the AP syllabi for chemistry, biology and physics**
  - This allows us to expand the types of reasoning addressed by Halo

- **Two competing teams/approaches (F-Logic, Concept Maps/KM)**

- **Evaluation and downselect in September 2006**

Team SRI Halo 2 Intermediate Evaluation

- Science grad student KBs
- Extensive natural lang
- ~$100 per syllabus page

Domain	Number of questions	Percentage correct			
		SME1	SME2	**Avg**	**KE**
Bio	146	52%	24%	**38%**	**51%**
Chem	86	42%	33%	**37.5%**	**40%**
Phy	131	16%	22%	**19%**	**21%**

VS.

Halo Pilot System	Percent correct
Cycorp	37%
SRI	44%
Ontoprise	47%

- Professional KE KBs
- No natural language
- ~$10K per syllabus page

Knowledge Formulation

- Time for KF
  - Concept: ~20 mins for all SMEs
  - Equation: ~70 s (Chem) to ~120 sec (Physics)
  - Table: ~10 mins (Chem)
  - Reaction: ~3.5 mins (Chem)
  - Constraint: 14s Bio; 88s (Chem)
- SME need for help
  - 68 requests over 480 person hours (33%/55%/12%) = 1/day

Question Formulation

- Avg time for SME to formulate a question
  - 2.5 min (Bio)
  - 4 min (Chem)
  - 6 min (Physics)
  - Avg 6 reformulation attempts
- Usability
  - SMEs requested no significant help
  - Pipelined errors dominated failure analysis

System Responsiveness

- Biology: 90% answer < 10 sec
- Chem: 60% answer < 10 sec
- Physics: 45% answer < 10 sec

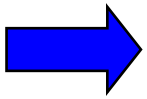
	Interpretation (Median/Max)	Answer (Median/Max)
Bio	3s / 601s	1s / 569s
Chem	7s / 493s	7s / 485s
Phy	34s / 429s	14s / 252s

No other system has EVER achieved this performance level with SME-entered knowledge

VULCAN

Talk Outline

- **The Halo Vision**
- **Systems AI – Vulcan's Halo Program**
  - The Halo Pilot: The Limits of Expert Systems
  - Halo Phase II: Deep Reasoning over the AP problem
  - Halo Today: Leveraging the Web
- **The Future of Halo**



VULCAN

The Halo Project Today

- SME Knowledge Entry and Question Answering Technology (Aura)
- Scaling up the KB (Offshore knowledge entry)
- SME entry and use of defaults and rule knowledge
- Scaling up Participation (Semantic Wikis)

VULCAN

Aura Goals for the September 2008 Evaluation

- **Demonstrate a 75% score for correctness and explanation on the intermediate evaluation questions, using SME authored KBs**
  - Current scores range from 16% to 52%
- **Median number of SME question reformulation attempts will be 5 or less (end-to-end)**
  - Current numbers are 5 (Chem); 7 (Physics); and 1 (Bio, constrained by limited possible question types)
- **Performance**
  - Complete 75% of the knowledge formulation operations in 5 sec or less
  - For 75% of the final evaluation questions, the mean response time for interpreting a question and answering a question will be less than 10 sec.
  - For 90% of the questions, the mean system response time for answering the question will be less than 1 minute

Scaling Up to the Digital Aristotle

- **Contracted KB construction**
  - Aura tested at IJCAI with IIIT-Hyderabad students
  - Investigating linkup with offshore institutions
    - Does the lower cost of student labor offset higher management costs?
    - Traditional consulting firms are too expensive
    - Looking for Indian bioinformatics firms
  - Next steps
    - Gather bids and select a performer
    - Pilot with the implementation phase syllabus (~160 hours); compare to reference and US results

- **Other options**
  - US-based student labor
  - Game-based knowledge acquisition

- **Goal: Pilot study ready to go by the time Aura is ready**

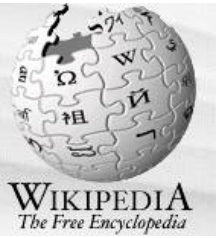
VULCAN

Halo and Rules Knowledge (New 2008 Thrust)

- **SILK: Suite of core knowledge representation and reasoning (KR) modules**
  - Provide defaults, hypotheticals, actions, and processes capabilities
    - First Focus: Combine defaults with as much as possible of other established features for monotonic (DB, classical, ontology). Default flavor pervades the KR
    - Key ideas: Courteous extension of Logic Programs, distributed, event-driven
    - Second Focus: Hypotheticals/Actions/Processes. Key ideas: advanced defaults and rules
  - Employ distributed algorithms and platform for high scalability
    - Focus: Incremental update/merge, with distributed dynamic import
    - Key ideas: dependency analysis, precomputation
  - Progressively/iteratively extend with new expressive features and algorithms
  - Early iterates, e.g., initial defaults, have substantial value for science and business/govt.
  - Interoperate via KR and SOA standards with other systems/sources, including web sources
- **Knowledge acquisition (KA) and UI modules, building on SILK KR**
  - Provide assert, query, answer, browse, edit, test, explain, analyze, debug capabilities
- **Integration of the above**
  - Into Aura, to significantly boost AP performance
  - Into Semantic MediaWiki (SMW) or other wiki/Web2.0 environment, for knowledge acquisition
  - As a stand-alone KR technology

Rethinking Halo in the Age of the Web

- **Halo 2's knowledge acquisition design is classic AI**
  - Halo systems (SRI, Ontoprise) are logically self-contained
  - Knowledge acquisition use cases are single-author expert systems



- **But, Vulcan's goal is the *Digital Aristotle***
  - Large knowledge bases in support of human inquiry
    - Scale beyond single authors to Web scale
  - Social issues surrounding real KR&R systems
    - Disciplinary approval of KB
    - Non-formal annotations of KB material (historical material, examples, different pedagogical approaches)
    - Transparency of motivation for KB modeling choices

YAHOO!

myspace.com a place for friends

flickr LOVES YOU™

Google™

- **So, we have made programmatic changes in Halo**
  - Expand knowledge acquisition approach
    - RDF/OWL import and export (for DL-expressible fragments)
    - Use Semantic Wikis (specifically, AIFB's Semantic MediaWiki)
    - Basic support for collaboration
  - Leverage European research vigor

YouTube Broadcast Yourself™

VULCAN

Semantic Wikis – The Main Idea

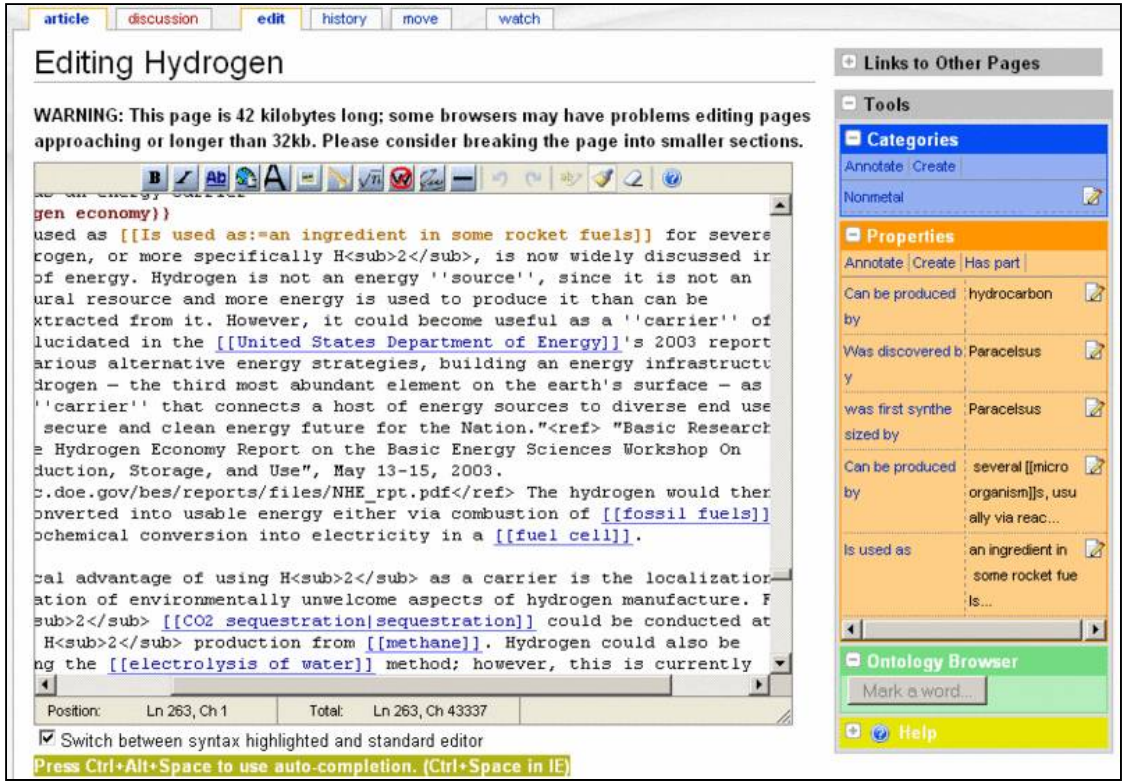
- **Wikis are tools for *Publication* and *Consensus***
- **MediaWiki (software for Wikipedia, Wikimedia, Wikinews, Wikibooks, etc.)**
  - Most successful Wiki software
    - High performance: 10K pages/sec served, scalability demonstrated
    - LAMP web server architecture, GPL license
  - Publication: simple distributed authoring model
    - Wikipedia: >2M articles, >180M edits, 750K media files, #8 most popular web site in October
  - Consensus achieved by global editing and rollback
    - Fixpoint hypothesis (2:1 discussion/content ratio), consensus is not static
    - Gardener/admin role for contentious cases
- **Semantic Wikis apply the wiki idea to basic (typically RDFS) structured information**
  - Authoring includes instances, data types, vocabularies, classes
  - Natural language text for explanations
  - Automatic list generation from structured data, basic analytics
  - Searching replaces category proliferation
  - Reuse of wiki knowledge

Semantic Wiki Hypotheses:

(1) Significant interesting non-RDBMS Semantic Data can be collected cheaply

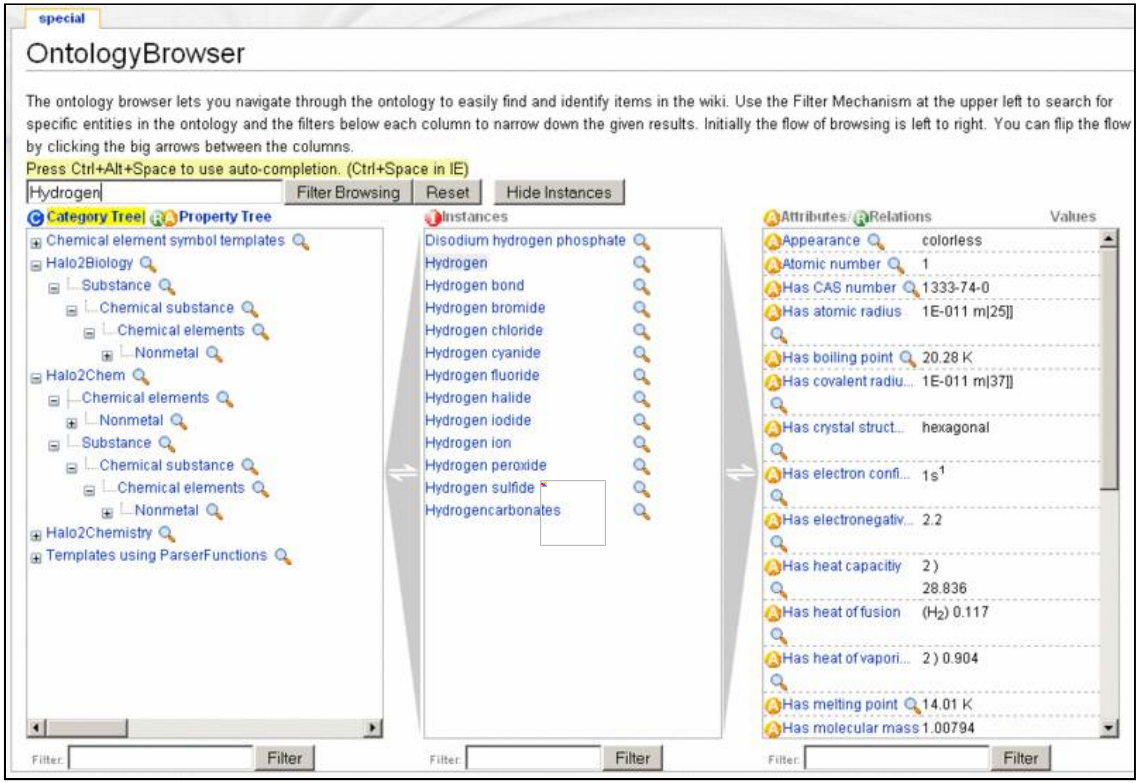
(2) Wiki mechanisms can be used to maintain consensus on vocabularies and classes

Semantic MediaWiki



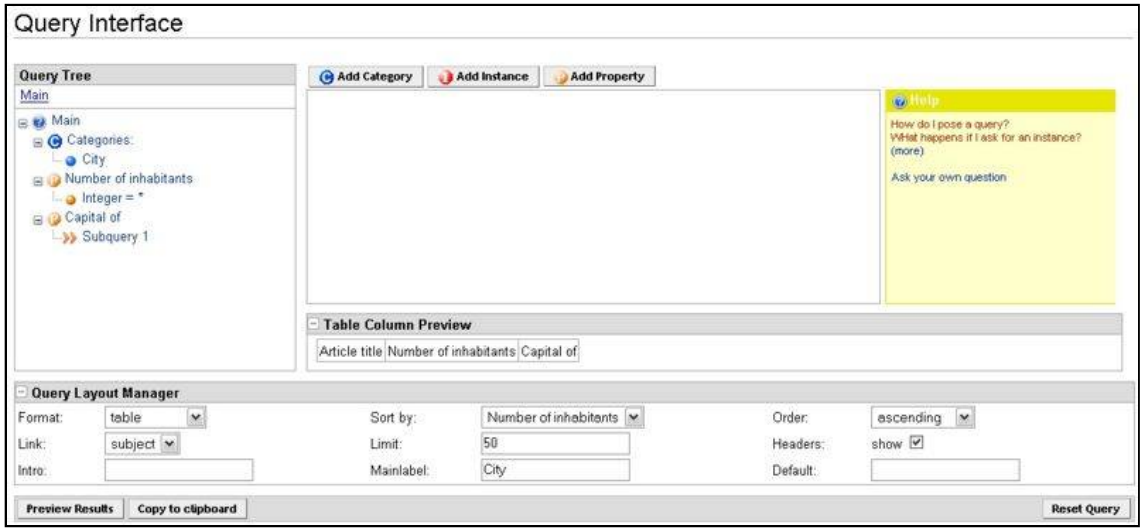
- **Knowledge Authoring Capabilities (SMW 1.0 plus Halo Extension)**
  - Syntax highlighting when editing a page
  - Semantic toolbar in edit mode
    - Displays annotations present on the page that is edited
    - Allows changing annotation values without locating the annotation in the wiki text
  - Autocompletion for all instances, properties, categories and templates
  - Increased expressivity through n-ary relations (available with the SMW 1.0 release)

Semantic MediaWiki



- **Semantic Navigation Capabilities (SMW 1.0 plus Halo Extension)**
  - GUI-based ontology browser, enables browsing of the wiki's taxonomy and lookup of instance and property information
  - Linklist in edit mode, enables quick access of pages that are within the context of the page being currently edited
  - Search input field with autocompletion, to prevent typing errors and give a fast overview of relevant content

Semantic MediaWiki

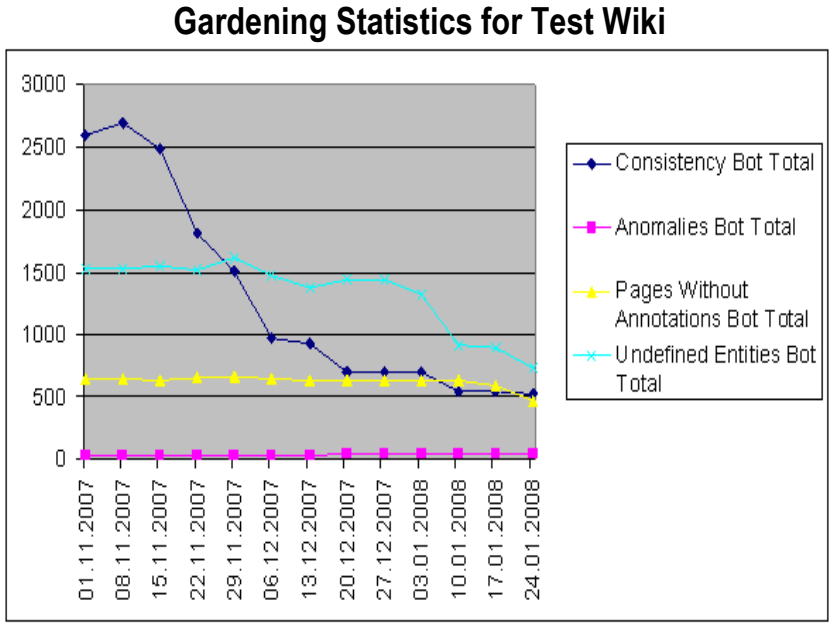


- **Knowledge Retrieval Capabilities (SMW 1.0 plus Halo Extension)**
  - Combined text-based and semantic search
  - Basic reasoning in ask queries with sub-/super-category/-property reasoning and resolution of redirects (equality reasoning)
  - GUI-based query formulation interface for intuitive assembly and output generation of ASK queries (no SQL/MQL/SPARQL)

- **Fully open source under GPL**
- **Extensive formal user testing**
- **Download at: http://semanticweb.org/wiki/Halo_Extension**

Cool Stuff... But Does it Work?

- **User tests were performed in Chemistry**
  - 20 graduate students were each paid for 20 hours (over 1 month) to collaborate on semantic annotation for chemistry
  - ~700 Wikipedia base articles
  - US high-school AP exams were provided as content guidance



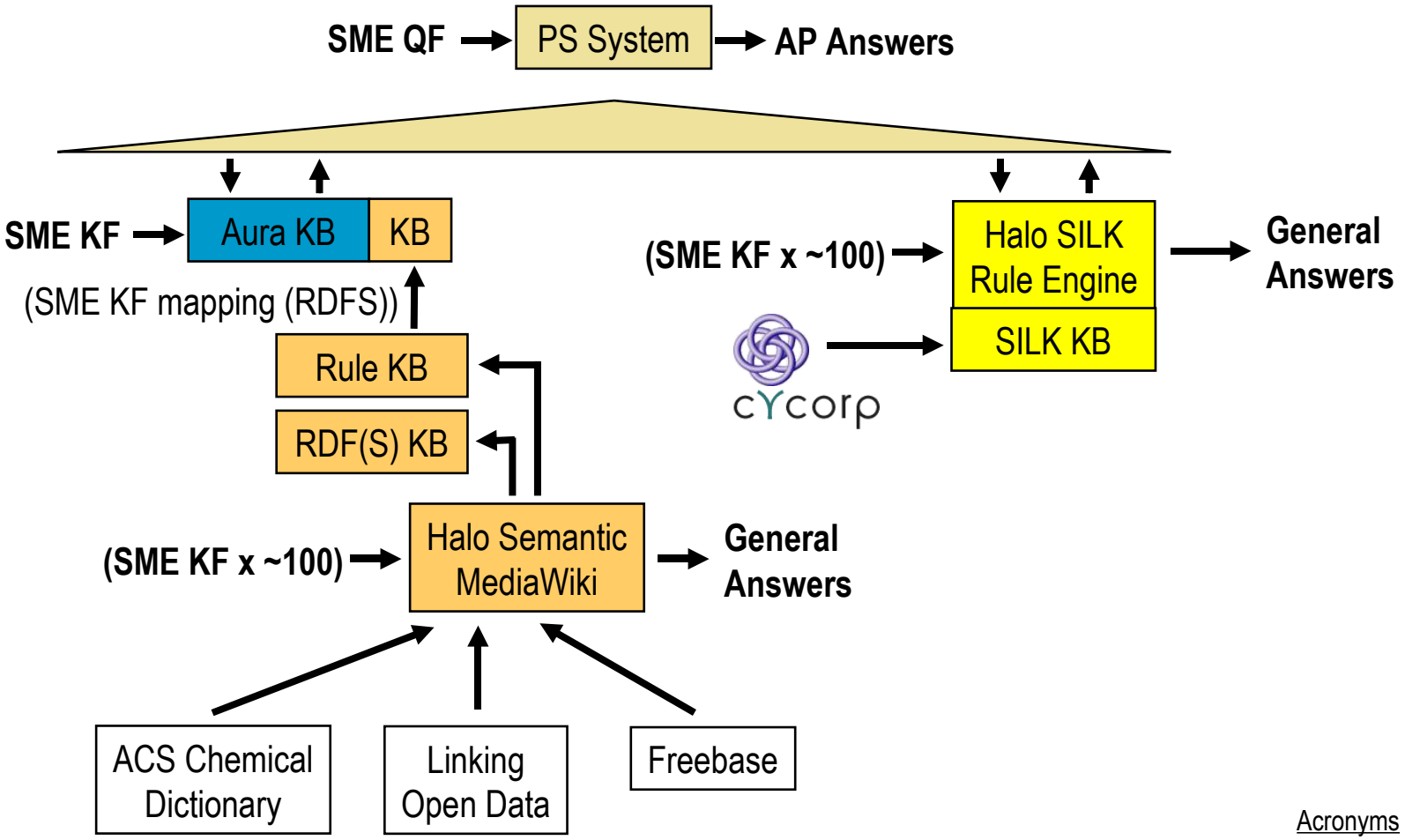
- **Initial Results**
  - Sparse: 1164 pages (entites), avg 5 assertions per entity
    - 226 Relations (1123 relation-statements) and 281 attributes (4721 attribute-statements)
  - Many bizarre attributes and relations
  - Very difficult to use with a reasoner

- **User testing and quality results for Phase II extensions**
  - Initial SUS scoring (6 SMEs, AP science task) went from 43 to 61; final scores in the 70s
  - 3 sessions using the Intrinsic Motivation Inventory (interest/value/usefulness); up 14%
  - Aided by the consistency bot, users corrected 2072 errors (80% of those found) over 3 months

- **Semantic Wikis for the Education Community**
  - "Everyman's Database" – blends text and data in a collaborative wiki environment
  - Scalable sharing tools with simple data analytics
  - Semantic wikis can redisplay data from other databases, and export data to other tools

VULCAN

Vulcan Project Halo Architecture



- Leverages Semantic Web for general knowledge
- Overlapping KRs and meta-level problem solving architectures
- Adds SILK for explicit rule formulation and reasoning

Acronyms
KB = Knowledge Base
KE = Knowledge Engineer
SME = Subject Matter Expert
KF = Knowledge Formulation
QF = Question Formulation
AP = Advanced Placement
PS = Problem Solving

Summary: Areas for Halo in 2008

Core Halo Thrust

- **Complete Phase II Aura and SMW**
  - 75% correctness with a 10 sec mean question answering time
  - Mean 5 user question reformulation attempts
  - Aggregate 30% reusable knowledge from SMW
- **Perform Halo Phase II Evaluation**
  - SME-driven KB construction
  - SMEs will pose AP-level questions
  - AP-level grading
  - SMW evaluation and wiki import
  - Phase III specific metric goals
- **Kickoff Halo Phase III Development**
- **Semantic Web Outreach**
  - Large Knowledge Collider (LarKC)
  - Networked Ontologies (NeOn)

Halo Advanced Research Thrust

- **Rule Engine Basic R&D**
  - More expressive/reasoning power
    - Focus on defaults, general rules, and formal processes
    - Hyper Logic Programs, Nonmonotonic, KR formalization of Aura
  - More commonsense knowledge (ReCyc)
    - Cyc knowledge used in SILK validation
    - If IP available, use of Cycorp reasoning modules
    - Cyc KB translation and import (goal 50%)
- **Rule Authoring R&D**
  - KA/UI prototypes for authoring SILK defaults in Halo
  - Simple Rules in SMW
- **Initial SILK Prototype**
  - Integration with Aura and SMW
  - Test with AP subject domains

Talk Outline

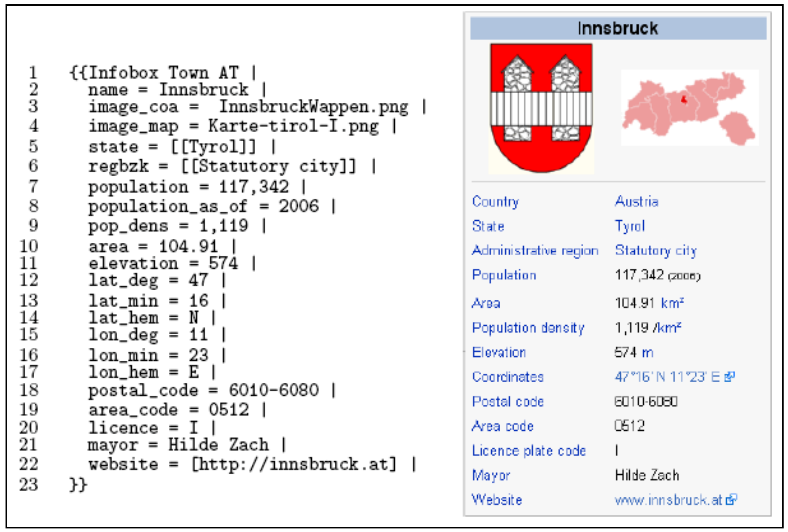
- **The Halo Vision**
- **Systems AI – Vulcan's Halo Program**
  - The Halo Pilot: The Limits of Expert Systems
  - Halo Phase II: Deep Reasoning over the AP problem
  - Halo Today: Leveraging the Web
- **The Future of Halo**

Core Halo: Building Outward

- **Build a more complete Halo**
  - Halo R&D targets the challenging problem of AI scale by SME-based authoring and deep question-answering technologies
    - We have had world-class success with our current approach
    - For a complete system, we need to combine Halo's unique technologies with other techniques (search, database-style query, approximate answers, etc.) that address technically easier problems
- **Build a Halo user community**
  - Halo Extensions for Semantic MediaWiki look like a solid success
    - Leverage open source community for software improvements and extensions
  - Tighter links to Semweb/Web 2.0 community
    - Sources of data, rules, commonsense knowledge
    - Sources of Knowledge for the Final Encyclopedia

A Knowledge Source for Halo: The DBpedia Project

- **Mine Wikipedia for assertions**
  - Scrape Wikipedia Factboxes
    - ~23M triples
  - High-confidence shallow English parsing
  - Category assertions
- **DBpedia 3.0 dataset**
  - ~2M things, ~220M triples
    - 80K persons, 293K places, 62K music albums, 36K films, 489K links to images, 2.7M links to relevant external web pages, 2.1M links into RDF datasets
  - Classifications via Wikipedia categories and WordNet synsets
  - One of the largest broad knowledge bases in the world
- **Simple queries over extracted data**
  - Public SPARQL endpoint
  - “Sitcoms set in NYC”
  - “Soccer players from team with stadium with >40000 seats, who were born in a country with more than 10M inhabitants”



```
{{Infobox Town AT |
  name = Innsbruck |
  image_coa =  InnsbruckWappen.png |
  image_map = Karte-tirol-I.png |
  state = [[Tyrol]] |
  regbzk = [[Statutory city]] |
  population = 117,342 |
  population_as_of = 2006 |
  pop_dens = 1,119 |
  area = 104.91 |
  elevation = 574 |
  lat_deg = 47 |
  lat_min = 16 |
  lat_hem = N |
  lon_deg = 11 |
  lon_min = 23 |
  lon_hem = E |
  postal_code = 6010-6080 |
  area_code = 0512 |
  licence = I |
  mayor = Hilde Zach |
  website = [http://innsbruck.at] |
}}
```


UNIVERSITÄT LEIPZIG

DBpedia

Freie Universität Berlin

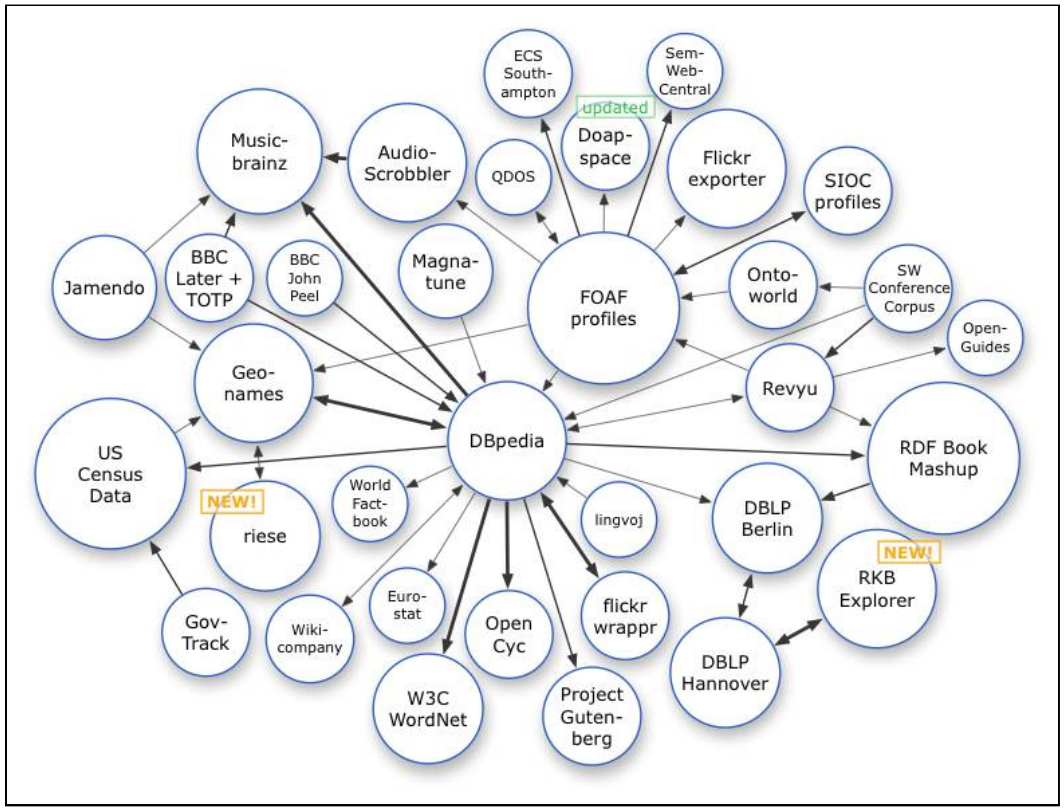
OPENLINK SOFTWARE

VULCAN

Linking Open Data

- **W3C Project primarily carried out in Europe**
- **Goals**
  - Create a single, simple access mechanism for web RDF data
  - Build a data commons by making open data sources available on the Web as RDF
  - Set RDF links between data items from different data sources
- **Total LOD dataset**
  - ~2B triples, and ~3B RDF links
  - Growing all the time (ex: 3B Eurostat triples)
  - Database linkage means that LOD will soon be impossible to count except via order of magnitude

LINKINGOPENDATA
W3C SWEO Community Project



VULCAN

Networked Ontology Project (NeOn)

- **Ever try to use 3-4 networked ontologies?**
  - Location and characterization of ontology resources
  - Version control under multiple revisions
  - SOA and mapping management
  - Lifecycle issues

- **NeOn is an EC Framework 6 Program (2006-2009)**
  - ~€15M, 14 partners including UN FAO, pharmaceutical distribution
  - Goals:
    - To create the first ever service-oriented, open infrastructure, and associated methodology
    - To support the overall development life-cycle of a new generation of large scale, complex, semantic applications
    - To handle multiple networked ontologies in a particular context, which are highly dynamic and constantly evolving.

- **Outputs: The open source (GPL) NeOn toolkit: http://www.neon-toolkit.org/**

VULCAN

Final Thoughts on Halo

- **Halo is one of the largest "classic AI" R&D programs in the US**
  - We bring together graduate students, research labs, and universities into an unified, ambitious project
  - Halo is known worldwide
- **Part of an increasingly-integrated strategy at Vulcan to invest in semantics and advanced knowledge tools**
  - Other investments: Radar Networks, ZoomInfo, Evri, Kiha, etc...
  - More in the pipeline
- **Semantic MediaWiki is a near-term spinout**

Thank You

VULCAN