Brittleness Taxonomy: Final Version

Purpose of Taxonomy

We are interested in identifying sources of brittleness in knowledge systems for two reasons. First, in the event that knowledge systems built for Halo fail to answer a chemistry question, or fail to explain the answer well, we would like to assess the reason for the failure. Second, we would like to better understand how the various vulnerabilities of knowledge systems affect their viability and application so that we might direct our future research efforts in an informed manner.

Definition of Brittleness

A brittle system is one that experiences a precipitous drop in its performance when it moves outside of its original scope of application. Systems can be brittle along different dimensions. And, being brittle along one dimension does not entail that a system will be brittle along another, though some of the categories below are related to one another. Scoring high on many different brittleness dimensions would warrant an overall assessment of 'brittle system', while scoring high in only one or two areas may warrant only an assessment of 'brittle in certain areas'.

VULCAN.COM

Description of Document

The classical view on brittleness is that it is caused primarily by the inability of a system to fall back on "common sense" or "first principles" knowledge. An expert system, for example, is so task oriented that its founding assumptions are not representable within the system itself; consequently, it cannot determine when a situation is out of its scope. This suggests that a non-brittle system would require encoding of "common-sense/first principles" knowledge in addition to task specific rules. Encoding large amounts of knowledge, however, introduces new kinds of brittleness. The taxonomy below is an attempt to capture some of the most salient examples.

505 Fifth Ave S Suite 900
Seattle, WA 98104

206 342 2000 Tel
206 342 3000 Fax

Key to Headings

1. **(MOD) Knowledge Modeling:** the ability of the knowledge engineer to model information/write axioms
2. **(IMP) Knowledge Implementation/Modeling Language:** the ability of the representation language to accurately represent axioms
3. **(INF) Inference and Reasoning:** the ability of the inference engine to "find the needle in the haystack"
4. **(KFL) Knowledge Formation and Learning:** the ability of the system (KB + inference engine) to acquire and merge knowledge through automated and semi-automated techniques
5. **(SCL) Scalability:** the ability of the KB to scale
6. **(MGT) Knowledge Management:** the ability of the system to maintain, track changes, test, organize, document; the ability of the knowledge engineer to search for knowledge
7. **(QMN) Query Management:** the ability of the system to robustly answer queries
8. **(ANJ) Answer Justification:** the ability of the system to provide justifications for answers in the correct context and resolution
9. **(QMT) Quality Metrics:** the ability of the developers to determine how "good" the knowledge base is at any given point in its evolution
10. **(MTA) Meta Capabilities:** the system's ability to utilize meta-reasoning or meta-knowledge

Brittleness	Influences	Description	Example	Mitigation	Future Research
B-MOD-1	MOD, QMT	**Modeling Error Brittleness** The knowledge engineer fails to capture domain information properly in their modeling (the act of writing the axiom).	Classifying chemical as an Acid independent of the reaction.	Review processes to validate that domain specific information is captured correctly; SME testing of the system; SME involvement throughout	Tools to better facilitate knowledge modeling by domain experts; Automated techniques to vet completeness and coverage of KB formation
B-MOD-2	MOD, MGT, INF, SCL	**Modeling Assumption Brittleness** Implicit "context" assumptions are not articulated, making it difficult for knowledge engineers to properly model/extend/modify information; Designers working from disparate assumed "context models" may introduce conflicts into the KB. Resolving multiple contexts may create large, unwieldy rule sets, difficulty in scaling	An explicit assumption is made about normal temperature. Unless the explicit assumption is seen, later OEs may introduce statements that conflict with the assumption.	Clearly document such important assumptions; Explicitly handle/represent "context" in the KB -- doing so will reduce the likelihood of such knowledge conflicts; use inference engine to alert OEs of such knowledge conflicts	Create a tool that will allow OEs to quickly search through the domain assumptions of a certain context; Have the KB alert the OE that domain assumptions relate to concepts they are using.
B-MOD-3	MOD, IMP	**Modeling Primitive Brittleness** The desired knowledge cannot be modeled in a straightforward manner using the modeling language.	Trying to model Fourier analysis in first-order logic.	Try to shoe-horn knowledge as best you can within the given framework, or use plug-in modules	Explore other reasoning capabilities, e.g. induction, abduction, etc..

Brittleness	Influences	Description	Example	Mitigation	Future Research
B-MOD-4	MOD	**"Islands of Knowledge" Brittleness** The knowledge engineer creates "islands of knowledge" by not making explicit connections between the domain being worked on and existing domains, in particular those which represent knowledge at a more general level.	Representing the relation between a mixture and the type of reaction occurring within it without making explicit the particular "in" relation that holds between the mixture and the reaction type.	Component Library or upper middle ontologies having good breadth and being able to easily integrate between the "islands" and the upper/middle ontologies; ontology mapping tools	Automated or semi-automated techniques to identify gaps in the middle ontology; better reuse techniques; attaching and distributed ontologies and inferencing
B-IMP-1	IMP, MTA	**Under-expressive Language Brittleness** The descriptive language used by the KB is incapable of representing concepts needed to capture the desired knowledge, thus approximations are used	First order logics cannot describe higher-order concepts like "belief": John believes that UFOs exist.	Utilize a more expressive language, or live with an incomplete description of the knowledge; design ways to better cope, e.g. plug-ins	What is the correct tradeoff between expressivity and complexity? Identify which representations are best for which tasks
B-IMP-2	IMP	**Over-expressive Language Brittleness** Using overly expressive language causes inference to be intractable	Using a highly expressive description logic takes a very long (or not finite) time to compute queries	Reducing expressiveness; sometimes the expressiveness is needed, so live with the consequences	Searching for adequate inferencing strategies, tractable subsets
B-IMP-3	IMP, MOD	**External Module Interface Brittleness** The descriptive language of the KB does not readily	Inability to access modules like Fourier Transforms, 3D modeling, etc. The transformation	Fit the external modules by using an expensive data structure	Hybrid reasoning architectures

Brittleness	Influences	Description	Example	Mitigation	Future Research
		translate to the representation states of external modules	between a sentence like "Fido is a dog" to the 3D depiction of Fido as a dog is extremely complex.	translation; Use sub-optimal methods for exchanging information.	
B-MGT-1	MGT, MOD	**Large KB Learning Brittleness** Large KBs are often difficult to learn because of the large number of concepts and the complex ways in which they relate to each other. Poor search and documentation tools compound this problem	An OE wanting to represent the relation of a solution to its component solutes must first search to see whether such a relation already exists or whether a more general relation might be preferable.	Search and browsing tools.	Indexing the correct concepts given meaning and context.
B-MGT-2	MGT, MOD, QMT, SCL	**Large KB Extension Brittleness** Large, highly interconnected KBs are difficult to extend correctly; The larger and more interconnected a KB gets, the more potential it has to introduce brittleness of type B-MOD-1, because the number of potential failures by omission increases with the magnitude and connection factor	To correctly ontologize the chemistry domain, the concept of a chemical compound needs to be represented. But if the concept of a compound has not been represented, then that needs to be done first. The amount of such background work can become prohibitive if the system is very brittle along the B-MOD-1 dimension.	Better quality control on knowledge formation and coverage; truth maintenance systems, schema techniques; better documentation	Automated techniques to vet completeness and coverage of KB formation.
B-MGT-3	MGT, MOD	**Large Team Brittleness**	A KB for a book is being built by dividing it amongst	Change management,	Better processes,

Brittleness	Influences	Description	Example	Mitigation	Future Research
		Collaboration /History of development/ Versioning; Large KBs will, invariably, be built using a large team. We need a way to coordinate their changes and to communicate any implicit assumptions that went into the modeling.	multiple OEs. A piece of knowledge is needed by multiple OEs, and they each encode their own version which needs to be reconciled. Knowledge entered in one chapter has consequences in other chapters. So, testing the knowledge involves checking the consistency of one OE's work, but also how it relates to others.	e.g. CVS and explicit assumption documentation; book-keeping assertions; explicit representation of workflow.	explicit representation of workflow.
B-KFL-1	KFL	**Information Extraction Brittleness** Due to exponential explosion of information, deriving automated techniques for extracting knowledge is required. Unfortunately, machine learning techniques are not well suited for extracting deep KRR from unstructured data.	System can't extract knowledge from books or the web without a human "in the loop".	Manual annotation of unstructured data, or using automated annotation tools.	Knowledge-based information extraction; human guided extraction.
B-KFL-2	MGT, KFL	**Knowledge Mapping Brittleness** With exponential explosion of information, it is critical to be able to merge structured	One DB may represent the title of a film as Film(film_name, film_title) while another might represent it as	Standards on KB formation and "exposed" APIs like those being developed for the semantic web.	Automated techniques to select optimal "exposure" points in an existing KB, possibly given

Brittleness	Influences	Description	Example	Mitigation	Future Research
		knowledge from multiple sources. Merging KBs is highly human labor intensive and error-prone. Current KB technologies provide no easy common ground to facilitate merging	(titleOfFilm FILM TITLE). Some independent source needs to make the equivalence of these two representations explicit before the DBs can be merged.		an objective; automated mapping techniques.
B-INF-1	INF, MOD	**Inference Engine Conceptualization Brittleness** User error in conceptualizing inferencing algorithms.	The default reasoning system caches the values and does no further reasoning if a value is locally available. The user fails to understand this, and expects the system to use rules to derive new values that are different from the cached values.	Tools to "visualize" the system's working to the OE/KE; explicit semantics available.	Explicit formalization of the knowledge; more user-friendly debugging tools.
B-INF-2	INF, IMP	**Inference Engine Bug Brittleness** Error in implementing the algorithm	The software implementation fails to properly capture the correct algorithm details	Employ proper software engineering, quality assurance procedures to minimize coding problems	Formal verification methods; symbolic level algorithm checkers.
B-INF-3	INF, MGT, SCL	**"Practical Incompleteness" Brittleness;** Deep KBs pose resource challenges that practically prevent exhaustive searches, thus potentially failing to return an answer despite the fact that	Some complex inferences involve very high degrees of backchaining.	Manual partitions; heuristically motivated searches.	Re-factoring of graphic representation in the KB could prevent knowledge to be "buried" too deeply, given reasonable initial

Brittleness	Influences	Description	Example	Mitigation	Future Research
		the information exists in the KB. In many instances searches may be highly sensitive to initial conditions, resulting in unpredictable results			conditions; semantic optimization techniques.
B-INF-4	INF, MOD	**Consistency Brittleness** Deductive reasoning systems that encounter hard contradictions fail; Given any hard contradiction, you can prove anything is true; Large KBs that encompass many topics can result in contradictions.	Given "All birds can fly," "A chicken is a bird," "A chicken cannot fly" you could prove that Al Gore is president	Break the KB into blocks that are internally consistent (microtheories); use truth maintenance to verify that hard contradictions do not exist; non-monotonic reasoning.	Explore other reasoning capabilities, e.g. induction, that do not have the same sensitivities; use different logics.
B-INF-5	INF	**Numeric Instability Brittleness** Lack of factoring numerical aspects of computation into query response formation may lead to incorrect answers.	In many sample questions, the computed values are often different from the correct answer values specified in the multiple-choice options.	Address issues of precision and accuracy.	Explore the ability to infer desired accuracy and precision from the "context" of the query.
B-QMN-1	QMN, INF	**Query Scoping Brittleness** Ignoring irrelevant information in queries; Some queries contain irrelevant or out-of-scope details, or domain specific implicit assumptions that must be recognized in order to successfully infer the	For example, sample question 35 specifies that a yellow precipitate forms during the reaction. Color of the precipitates is not in the curriculum scope, but it is also unnecessary to solve the problem.	Some irrelevant facts can be ignored by the inferencing engine, allowing the task to still be completed.	Explore more systematic approaches to inferring implicit "context" and determining query scope.

Brittleness	Influences	Description	Example	Mitigation	Future Research
		query's answer.			
B-QMN-2	QMN, MOD, INF	**Query Encoding Brittleness** Sensitivity to query encoding.	Slight manipulation of the query encoding results in dramatic changes in the system's ability to correctly answer the question.	Intensive testing against known query types to guarantee performance.	Better systemic understanding of encoding sensitivity, better testing tools; semantic query optimization technique.
B-ANJ-1	ANJ, INF, QMT	**Exposition Brittleness** Inference proof trees are inadequate by themselves for constructing human understandable answer justifications; Humans require answer justifications to gain confidence that the system is working properly. Proof trees contain lots of irrelevant and out of sequence information, which also might be at an inappropriate level of specificity. Desired resolution of the justification is also context and user specific.	If we say "5.02 = 5.00 +/- 5%" then this is clear (or at least believable) to a human, but is many steps from being formally proven.	Designing KBs with concepts that take into account the need to justify answers can create more legible proof trees; post processing on proof trees.	Deriving justification context that is informed by the user, topic and line of questioning; exploring middle ontologies for aid in presenting justifications; post-processing on proof trees.
B-ANJ-2	ANJ, IMP, MGT	**Answer Template Brittleness** Limitations of manually created answer justification templates;	If a new rule, which allows a number of questions to be answered in a different, perhaps more straightforward	Maintaining strict revision control and quality assurance on manually formed	Migrating to automatically formed answer justifications; moving towards

Brittleness	Influences	Description	Example	Mitigation	Future Research
		Manually constructed answer templates must be constantly updated to reflect evolving contexts. Difficulty in supporting multiple resolutions.	manner, is introduced into the system, all of the answer templates for the affected questions will need to be manually changed accordingly.	rules.	fully automated justification generation; intermediate explanation representation.
B-ANJ-3	ANJ	**Context Justification Brittleness** Inability to produce user and context appropriate justifications.	Justifying answers to AP chemistry students might be very different from those provided to naïve users.	Relying on context sensitive mechanisms of the query context to determine the degree of detail that should be entailed in the justifications.	A systemic approach that is flexible enough to infer the level of detail required from an answer justification, given a number of input parameters.
B-QMT-1	QMT	**Quality Metrics Brittleness** Degree to which metrics are used to track and measure KB quality in terms of its coverage of the material and its ability to answer questions; KRR work proceeds without good metrics to indicate whether the effort is "converging", how much of the space is "covered" or when the work is "finished".	An OE adds a piece of knowledge but does not record this fact or test its performance in any way.	Exposing the system to many different question types and collecting performance information, including question encoding during the OE process.	Automated techniques to vet completeness and coverage of KB formation
B-MTA-1	MTA	**Meta Capabilities Brittleness** The system's inability to do meta-reasoning	A system may have knowledge of Chemistry, but not knowledge of	Some meta-knowledge can be represented with the same	Investigate language extensions to make meta-

Brittleness	Influences	Description	Example	Mitigation	Future Research
		or its lack of meta-knowledge. A system may be able to reason about its knowledge, but not able to reason about its reasoning.	Chemistry knowledge. For example, a system may have a law encoded that can compute a chemical's pH based on the log of [H+]. But it may not know why it's valid to do such a computation, or why pH is of interest to Chemists, etc.	KR language as regular knowledge (with appropriate modeling conventions and additions to primitives). Meta-reasoning may be beyond the capabilities of a given KRL.	knowledge capture as simple as knowledge capture. Investigate inference engine extensions to allow queries about reasoning in addition to queries about knowledge.