

e-Rulemaking: Research Problems for IT

This document attempts to synthesize some of the discussions of the e-Rulemaking workshop at KSG in January 2003. The perspective and organization taken here is explicitly chosen to support a report and an eventual call for proposals by the NSF.

It became obvious in the workshop that the timeline of the e-rulemaking process provides a natural organization for the IT research needs in this arena. That is therefore the organization adopted here:

1. Assembly and analysis: locating stakeholders and acquiring background material
2. Creating rules
3. Capturing commentary
4. Analyzing rules and commentary
5. Post-promulgation information collection and archiving
6. Usage and effectiveness of rules
7. Overall administration of rule creation and subsequent lifetime

With each point we describe the issue, provide one or more research questions, phrased so as to support experiments measuring the effect of IT to be developed, and then the types of IT most relevant to the issues.

1. Assembly and analysis: locating stakeholders and acquiring background material

Problem: In early stages of rule writing, rule-writers collect and systematize all kinds of information in order to justify their work. This source info may be studies commissioned by the Agency, existing studies performed by others, existing rules created earlier by the Agency and others, articles in the press, statements by pressure groups, letters from the public, etc. In addition to collecting information, rule-writers may decide to search for interested parties proactively, in order to bring citizens into the process early.

Research question 1: How much does IT help rule-writers find more and/or better background material required for writing the rule? Evaluation metrics: speed of acquisition, completeness/coverage of material, quality of rule justifications, etc.

Research question 2: How well does IT help rule-writers find appropriate stakeholders? Evaluation metrics: number and appropriateness of stakeholders

Relevant IT:

a: IR, database wrapping and access planners, ontologies, data mining, presentation and display

b: IR, information extraction, summarization, text mining, NLP semantic analysis, clustering, interfaces

c: IR, privacy maintenance

a. IT Research challenges, for numerical data

- Location: find relevant data from the world
 - Acquisition: incorporate and align heterogeneous data collections
 - Dispersal: cluster and stream apart data from different sources
 - Update: recognize and manage significant changes in source data and data collection processes
 - Analysis: identification of relevant data, and conversion to appropriate form
- b. IT Research challenges, for textual information**
- Location: find relevant text from the world
 - Acquisition and analysis: extraction of relevant material, categorization and clustering, automated recognition and derivation of internal and crosswise relationships, esp to rule under construction
- c. IT Research challenges, for locating stakeholders**
- Privacy: some people may not want others to know they are playing a role

2. Creating rules

Problems: Rules can be dense (therefore hard to understand) and inconsistent (both internally and with respect to other rules). This structure may be simply a partitioning of rule into its portions as separately annotated text blocks. More complex, it may include style-checking software that reminds or requires the rule-writer to create certain kinds of text in certain portions (for example, question syntax in the preamble, bullet points for each requirement specification, imperative sentences for the requirements, etc. It may limit the rule-writer to use certain terminology, or to provide hyperlinks to associated documents, etc. Structured rules may provide a number of benefits:

- pointed commentary (see below, topic **),
- automated cross-reference with background material, other rules, etc.,
- consistency checking,
- after rule promulgation, rule usage and enforcement.

Also, during rule-writing, writers may need to think through various alternatives. IT can help them model and simulate scenarios and record the results as justification for their eventual rules.

Research question 1: How useful would IT be in helping the rule-writer produce a suitably structured rule (i.e., does the structure enable more benefits than the work it takes)? Metrics: creation time vs time for subsequent processing

Research question 2: What degree of structure would rule-writers accept, and would IT be flexible enough to accommodate different degrees of variation for different rules?

Metrics: rule-writer questionnaire

Research question 3: Can general simulation/modeling IT packages be built that are still specific enough to be pertinent and useful? Metrics: utility, generality, cost, ease of use

Relevant IT: information extraction, clustering, similarity judgment IR/NLP, simulation, modeling, visualization tools

IT Research challenges:

- Rule templates and interfaces, internal format correctness checking
- Cross-indexing/similarity measurement; document structure analysis
- Rule material consistency checkers: semantic analysis, argument dependency graphs
- Scenario / modeling / simulation tools
- Visualization tools

3. Capturing commentary

Problems: The commentary request and collection stage may engender a large response, much of which semi-duplicated, some of which containing little of substantive value. IT may help focus commentary, structure it, relate it to the rule in consideration, and highlight possibly valuable material. In addition, the rule-writer may need to initiate an ongoing, multi-stage deliberation; IT may help guide and summarize this.

Research question 1: How can IT facilitate the treatment of commentary by rule-writers? In particular, to build commentary input websites that guide the citizens to focus their comments, to provide supporting material, etc. Also, IT can perform clustering (by similarity of content, of opinion, of specific topic, etc.), commentary summarization or highlighting, etc. Metrics: speed, coverage, completeness of rule-writers' treatment of commentary

Research question 2: How can IT best support ongoing multi-stage dialogue? Metrics: rule-writer questionnaire

Relevant IT: text classification according to various criteria, summarization, other NLP, bboard management, dependency maintenance systems, interfaces

IT Research challenges:

- tools to automatically create a commentary questionnaire / website with appropriate subsections
- technology used in Frequently Asked Questions classification, used by large companies to automatically answer most online user questions, can help focus citizens and provide pre-anticipated responses to common comments
- NLP tools to automatically analyze and interpret commentary, to cluster it by genre, opinion, source, detail, etc.
- technology to support various levels of dialogue, from chat rooms to distance meetings, records all the discussion, and cross-links it into the rules as appropriate

4. Analyzing rules and commentary

Problems: Rules and commentary can be more effectively created and managed if their component parts are cross-linked in various ways: from comment into rule, from rule into

supporting material, from comment to supporting material, from rule, supporting material, and comment to other, pre-existing rules (also from other Agencies), etc. Any new relevant material found automatically should be brought to the rule-writer's attention. If possible, such analysis systems could check for inconsistencies among portions of rules and/or material, and bring any inconsistencies to the rule-writer's attention.

Research question 1: Can IT use the structure of a rule and/or commentary to perform automated cross-indexing of rules, or link rules to relevant data, text, external services, etc. automatically? Metrics: degree and correctness of cross-indexing; satisfaction of citizens in locating all material of interest to them

Research question 2: Can IT perform automated consistency checking? Metrics: degree and correctness of consistency problems found (and missed)

Relevant IT: information extraction, clustering, similarity judgment IR/NLP, visualization tools

IT Research challenges:

- Rule templates and interfaces, internal format correctness checking
- Cross-indexing/similarity measurement; document structure analysis
- Rule material consistency checkers: semantic analysis, argument dependency graphs
- Scenario / modeling / simulation tools
- Visualization tools

5. Post-promulgation (publication) information collection and archiving

Problems: Once the rule is promulgated, IT can help disseminate it, support its use, track its life, etc. Dissemination includes converting rules and commentary to non-mainstream groups, e.g., the visually and otherwise handicapped, and people speaking other languages. Supporting use includes helping people (enforcement agencies as well as citizens with plans) judge the conformance of their enterprises. IT can help locate and record all subsequent 'activity' around the rule: press notices, court cases, etc.

Research question 1: What IT makes the information available best, to all stakeholders: enforcement, the disabled, non-English speakers, etc.? How well does this work? Metrics: various

Research question 2: What IT is useful for the continuous collection and integration of relevant material through the ongoing lifetime of the rule? How successful is it? Can this IT cross-link the new information to the most relevant portions of rules and background material? Metrics: systems, coverage, utility, correctness

Relevant IT: MT systems, handicapped aids, IR, classification, interfaces

IT Research challenges:

- technology to assist the handicapped with regard to rules
- machine translation of rules
- systems that automatically locate in the press and elsewhere any information relevant to the rule and record it

6. Usage and effectiveness of rules

Problems: IT can assist enforcers and the public in actually using the rule throughout its life. Rule conformance software that indicate to people with plans which parts of their plans need to be changed, and how, and systems that automatically locate all rules pertinent to their plans can be of great use to the general public. In addition, IT may be able to help evaluate the success of a rule: what do people say about it? How many fines are issued?

Research question 1: How well can IT ‘understand’ a citizen’s plans in order to locate all relevant rules? Metrics: recall and precision

Research question 2: How well can IT ‘understand’ a citizen’s plans and compare them to a (possibly structured) rule, in order to indicate rule compliance? How can compliance, and non-compliance, be best displayed and explained? Metrics: questionnaires for citizens, # of successful compliance experiences

Research question 3: How can you evaluate a rule? How well can IT use rule structure to help citizens determine whether their plans conform? Metrics: satisfaction of rule writers

Relevant IT: IR, NLP, semantic modeling and dependency analysis, QA, evaluation/measure tools: data capture, analysis, and display

IT Research challenges:

- technology to interpret citizen plans semantically
- technology to compare rule requirements and user plans, and locate differences
- technology to explain and display reasoning chains
- evaluation metrics 1 (performance): # comments, diversity, quality, cost per X, time per X
- evaluation metrics 2 (outcome): satisfaction, litigation, legitimacy, social results

7. Overall administration of rule creation and subsequent lifetime

Problem: Managers of rule-writers and archivists can use a single system that shepherds the rule throughout its whole lifetime, from inception, through promulgation, to eventually being superseded. Example information to be recorded: the statute, the authors, commentators, amounts of time spent in each stage, ancillary documentation,

ongoing role in the world (as recorded in the press, lawsuits, etc.), etc. All this to be placed into an archival record.

Research question: What is the most useful configuration of modules to support rule-writing and ongoing archiving? Metrics: software assembly, questionnaires for Gov managers

Relevant IT: scheduling software, versioning software, digital libraries archiving software, software packaging

IT Research challenges:

- bookkeeping system, with tools for noticing late schedule, bottlenecks, reports, etc.
- archiving support

Authors:

- Eduard Hovy, USC/ISI
- January 2003