Natural Language Processing Support for eRulemaking

Claire Cardie (PI)  Computing & Information Science
Cynthia Farina (PI)  Law School
Tom Bruce  Legal Information Institute
Erica Wagner  Hotel School

Funded (NSF) in November, 2005
Research Objectives

- Computer science
  - use well-developed and emerging methods in natural language processing (NLP) to develop tools to aid agency rule writers in:
    - organizing, analyzing, and managing the comments, studies, and other supporting documents associated with a proposed rule
    - analyzing proposed rules to flag possibly relevant legal mandates
Research Objectives

- Computer science
  - use well-developed and emerging methods of natural language processing (NLP) to develop tools to aid agency rule writers in:
    - organizing, analyzing, and managing the comments, studies, and other supporting documents associated with a proposed rule
    - analyzing proposed rules to flag possibly relevant legal mandates
Research Objectives

- Law
  - study the effect of technology on the rulemaking process
    - add to scarce existing information on actual rulemaking practice within agencies
    - identify effects of ongoing conversion to e-rulemaking on process and content
    - determine nature and incidence of common elements across wide range of agencies and diverse types of rules
    - work toward developing a set of “best practices for e-rulewriters”
Research Objectives

- Social science
  - apply qualitative and quantitative methods from social sciences to
    - understand the work- and information flow of the rulewriting process
    - guide system development to best support agency rulewriters
    - evaluate the integration of the NLP tools into, and their effect on, the day-to-day rulemaking process
Collaboration

- Department of Transportation (Neil Eisner)
  - Access to data
  - Access to rulewriters, legal staff, and analysts for surveys, ethnographic studies, usability studies

- Shulman/Callan/Hovy DG project
  - Data sharing
  - Using Callan’s Lemur IR system
  - Coordination on tool development
# Issue-channeled Solicitation for Categorization of Comments

## Document Search Results

<table>
<thead>
<tr>
<th>Docket ID</th>
<th>Document ID</th>
<th>Document Title</th>
<th>Date Posted</th>
<th>Type</th>
<th>Views</th>
<th>Add Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMS-2005-0049</td>
<td>AMS-2005-0049-0001</td>
<td>Marketing Order Regulating the Handling of Spearmint Oil Produced in the Far West; Revision of the Salable Quantity and Allotment Percentage for Class 3 (Native) Spearmint Oil for the 2005-2006 Marketing Year</td>
<td>12/05/2005</td>
<td>RULES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMS-2006-0004</td>
<td>AMS-2006-0004-0001</td>
<td>Hazelnuts Grown in Oregon and Washington; Establishment of Final Free and Restricted Percentages for the 2005-2006 Marketing Year</td>
<td>01/12/2006</td>
<td>RULES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMS-2006-0015</td>
<td>AMS-2006-0015-0001</td>
<td>Regulations Under the Perishable Agricultural Commodities Act (PACA)</td>
<td>01/30/2006</td>
<td>PROPOSED RULES</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMS-2006-0017</td>
<td>AMS-2006-0017-0001</td>
<td>Oranges, Grapefruit, Tangerines, and Tangelos Grown in Florida; Increased</td>
<td>02/01/2006</td>
<td>RULES</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Issue-channeled Solicitation for Categorization of Comments
Issue-channeled Solicitation for Categorization of Comments

Advance Notice of Proposed Rulemaking to Designate Critical Habitat for 20 Listed Evolutionarily Significant Units of Pacific Salmon and Steelhead

Closing Date - 11/13/2003 05:00 PM EST

View Proposed Rule
View Proposed Rule Summary

NOAA Fisheries has summarized the following issues for consideration in compiling your comments:
- Areas & Features That May Qualify as Critical Habitat
- Assessing the Benefits and Impacts of Designating Particular Areas
- Peer Review

You may comment on any of the issues or [Comment on the Whole Rule].

Each comment is limited to 4000 characters, but you may submit multiple comments.

List of Issues
- Proposed Rule Summary
  - Areas & Features That May Qualify as Critical Habitat
  - Assessing the Benefits and Impacts of Designating Particular Areas

Proposed Rule Summary

SUMMARY: NOAA Fisheries will be preparing critical habitat designation proposals for five species of Pacific salmon and steelhead (Oncorhynchus spp.) listed under the Endangered Species Act (ESA). The designations will address 20 evolutionarily significant units (ESUs) of these species in the states of WA, OR, ID, and CA. This advance notice of proposed rulemaking identifies issues for consideration and evaluation, and solicits comments regarding these issues as well as information regarding the areas and species under consideration.
Proposed Rule Summary

SUMMARY: NOAA Fisheries will be preparing critical habitat designation proposals for five species of Pacific salmon and steelhead (Oncorhynchus spp.) listed under the Endangered Species Act (ESA). The designations will address 20 evolutionarily significant units (ESUs) of these species in the states of WA, OR, ID, and CA. This advance notice of proposed rulemaking identifies issues for consideration and evaluation, and solicits comments regarding these issues as well as information regarding the areas and species under consideration.

DATES: Comments and information regarding the suggested designation process and areas being considered for designation may be sent to the appropriate address or fax number (See ADDRESSES), no later than 5 p.m. on November 13, 2003.

ADDRESSES: In addition to this e-comment site, comments may be sent to Chief, Protected Resources Division, NMFS, 525 NE Oregon Street - Suite 500, Portland, OR 97232. Comments may also be sent via facsimile (fax) to 503 230-5435.

FOR FURTHER INFORMATION CONTACT: Steve Stone, NMFS Northwest Region (WA, OR, and ID), 503/231-2317; Craig Wingert, NMFS Southwest Region (CA), 562/980-4021; or Lamont Jackson, NMFS Office of Protected Resources, Silver Spring, MD, 301/713-1401.

Comment on the Whole Rule

Back to Top

Areas & Features That May Qualify as Critical Habitat

The ESA defines critical habitat under section 3(5)(A) as "(i) the specific areas within the geographical area occupied by the species, at the time it is listed . . . , on which are found those
Areas & Features That May Qualify as Critical Habitat

The ESA defines critical habitat under section 3(5)(A) as: "(i) the specific areas within the geographical area occupied by the species, at the time it is listed . . ., on which are found those physical or biological features (I) essential to the conservation of the species and (ii) which may require special management considerations or protection; and (ii) specific areas outside the geographical area occupied by the species at the time it is listed . . . upon a determination by the Secretary that such areas are essential for the conservation of the species."

Section 3(5)(C) of the ESA goes on to state: "Except in those circumstances determined by the Secretary, critical habitat shall not include the entire geographic area which can be occupied by the threatened or endangered species."

Issues identified in this Notice as related to the ESA definition of critical habitat and for which the agency is soliciting comment include the following:

- What areas were occupied by the species at the time of listing? (Note Maps may be mailed to: Chief, Protected Resources Division, NMFS, 525 NE Oregon Street, Suite 500, Portland, Oregon, 97232)
- Are areas outside those currently occupied “essential for conservation”?
- Should the approach to critical habitat designation for hatchery fish be the same as for naturally spawned fish and if not, what approach should be used?
- What physical and biological features are essential to the species’ conservation?
- Are those essential features ones that may require special management considerations or protection and if so, what are those considerations/protentions?

Compose and Submit Comment on the Areas and Features That May Qualify as Critical Habitat
Initial Results

- Data set: only 2 rules
  - Protocol for Access to Tissue Specimen Samples from the National Marine Mammal Tissue Bank, Closing Date - 12/12/2002 05:00 PM EST
  - Alaska Groundfish Fisheries Draft Programmatic SEIS, Closing Date - 11/06/2003 11:59 PM EST
- Presumes campaign comments are non-existent or previously detected
- SVM’s, 5-fold cross-validation
- Training: issue-channeled comments, text of issues, relevant sections of rule (manually identified)
- Results
  - ~96% accuracy
Reducing Manual Annotation Costs

- Active learning

Lewis & Catlett [1994]
Active Learning Study

- **Data**
  - USDA-03-080-1 regarding Bovine Spongiform Encephalopathy
  - Identified 9 issues
  - 1676 comments
  - annotated 83 comments

- **Results: Mixed**
  - Unlabeled data set too small, not diverse enough
Accuracy Comparisons

![Graph showing accuracy comparisons between active and random selection methods.](image)

- Blue line: ALAlss1
- Pink line: RLAiss1
- Yellow line: ALAlss3
- Red line: RLAiss3
- Orange line: ALAlss7
- Black line: RLAiss7

**Active**

**Random**
## Friendlier Presentation of Proposed Rule

<table>
<thead>
<tr>
<th>Docket ID</th>
<th>Document ID</th>
<th>Document Title</th>
<th>Date Posted</th>
<th>Type</th>
<th>Views</th>
<th>Add Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMS-2005-0049</td>
<td>AMS-2005-0049-0001</td>
<td>Marketing Order Regulating the Handling of Spearmint Oil Produced in the Far West; Revision of the Salable Quantity and Allotment Percentage for Class 3 (Native) Spearmint Oil for the 2005-2006 Marketing Year</td>
<td>12/05/2005 RULES</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMS-2006-0004</td>
<td>AMS-2006-0004-0001</td>
<td>Hazelnuts Grown in Oregon and Washington; Establishment of Final Free and Restricted Percentages for the 2005-2006 Marketing Year</td>
<td>01/12/2006 RULES</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMS-2006-0015</td>
<td>AMS-2006-0015-0001</td>
<td>Regulations Under the Perishable Agricultural Commodities Act (PACA)</td>
<td>01/30/2006 PROPOSED RULES</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMS-2006-0017</td>
<td>AMS-2006-0017-0001</td>
<td>Oranges, Grapefruit, Tangerines, and Tangos Grown in Florida; Increased</td>
<td>02/01/2006 RULES</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
From the Federal Register Online via GPO Access [wais.access.gpo.gov]
[DOCID:fr10ap06-3]

DEPARTMENT OF AGRICULTURE

Agricultural Marketing Service

7 CFR Part 923

[Docket No. FV06-923-1 IFR]

Sweet Cherries Grown in Designated Counties in Washington: Removal of Container Regulations

AGENCY: Agricultural Marketing Service, USDA.

...
ACTION: Interim final rule with request for comments.

SUMMARY: This rule removes the container regulations prescribed under the Washington sweet cherry marketing order. Specifically, this rule removes the requirement that dark-colored sweet cherries must be handled in containers having a certain net weight. The marketing order regulates the handling of fresh sweet cherries grown in designated counties in the State of Washington, and is administered locally by the Washington Cherry Marketing Committee (Committee). By eliminating the container requirements, this relaxation will provide handlers with the ability to meet the rapidly changing wholesale, retail, and consumer demand for innovative product packaging. This is expected to enhance industry marketing flexibility and efficiency.

DATES: Effective April 11, 2006. Comments received by June 9, 2006 will be considered prior to issuance of a final rule.

ADDRESSES: Interested persons are invited to submit written comments concerning this rule. Comments must be sent to the Docket Clerk, Marketing Order Administration Branch, Fruit and Vegetable Programs, AMS USDA 1400 Independence Avenue SW STOP 0237 Washington, DC
SUPPLEMENTARY INFORMATION: This rule is issued under Marketing Agreement and Order No. 923 (7 CFR part 923) regulating the handling of sweet cherries grown in designated counties in Washington, hereinafter referred to as the "order." The order is effective under the Agricultural Marketing Agreement Act of 1937, as amended (7 U.S.C. 601-674), hereinafter referred to as the "Act."

The Department of Agriculture (USDA) is issuing this rule in conformance with Executive Order 12866.

This rule has been reviewed under Executive Order 12988, Civil Justice Reform. This rule is not intended to have retroactive effect. This rule will not preempt any State or local laws, regulations, or policies, unless they present an irreconcilable conflict with this rule.

The Act provides that administrative proceedings must be exhausted before parties may file suit in court. Under section 608c(15)(A) of the Act, any handler subject to an order may file with USDA a petition stating that the order, any provision of the order, or any obligation imposed in connection with the order is not in accordance with law and request a modification of the order or to be exempted therefrom. Such
Linking Rule Segments to Supplementary Information

- **Input**: proposed rule from Federal Register
- **Method**: standard ad-hoc IR
  - identify sections (e.g. agency, action, summary, dates, background) and sentences
  - for each sentence of rule, identify *all* or *the most* relevant sentences of background (Lemur)

- **Results**
  - 70% of rule sentences linked to relevant background segment
  - 6 rules
Related Project

- Spawned a related project to assess usability of existing web interface for e-rulemaking and to conduct expert redesign effort
  - Topic of workshop white paper
Collaboration Examples

- Obtained DOT data gathered by Jamie Callan (CMU)
  - Used in one study
- Determined key upcoming DOT and Commerce rulemakings suitable for observation and potential data sets
  - May (DOT): JARC/New Freedom/5310 Circular (public transportation)
  - June (NMFS): Salmon Critical Habitat Re-Designation (Seattle/Portland)
  - July (DOT): Transportation of Individuals with Disabilities
  - August (NMFS): NWHI Coral Reef Ecosystem Reserve designation as a national marine sanctuary
- Pending offer from Stuart Shulman for data annotation
Broad Impact

- NLP methods are general-purpose
  - useful in any context that requires managing, organizing, and analyzing large volumes of text
  - techniques that help agency rule writers can be used to design agency websites that help the public search, sort, and otherwise selectively access materials in the rulemaking process
  - E-Government Act of 2002 goal of making rulemaking “more transparent and accountable” and more “citizencentric.”

- all data sets and tools will be made publically available

- unique interdisciplinary opportunity for graduate and undergraduate students in computer science, law, and social science to investigate one of the most important public policy-making processes in contemporary society
Research Value of DG Domain

- Potential to advance the state-of-the-art in NLP
  - new data \(\rightarrow\) new issues for NLP \(\rightarrow\) new algorithms
  - sentiment analysis
Recommendations for Improving DG Program

- Hmmm...maybe we’ll have something to say regarding this next year
**Title**
Importation of Baby Corn and Baby Carrots From Zambia

We are proposing to amend the fruits and vegetables regulations to allow the importation into the continental United States of fresh, dehusked immature (baby) sweet corn and fresh baby carrots from Zambia. As a condition of entry, both commodities would be subject to inspection at the port of first arrival and would have to be accompanied by a phytosanitary certificate with an additional declaration stating that the commodity has been inspected and found free of the quarantine pest listed on the certificate. This action would allow for the importation of Zambian baby corn and baby carrots into the United States while continuing to provide protection against the introduction of quarantine pests.

**Abstract**

**Type**
PROPOSED RULES

**Sub-Type**
Proposed rule

**CFR Citation**
7 CFR 319

**Effective Date**

**Paperwork Control No.**

**RIN**

**Federal Register Number**
E6-00134

**Date Posted**
01/11/2006

**Comment Start Date**
01/11/2006

**Comments Due**
03/13/2006

**Compliance Date**

---

---

---