



Research on English and Foreign Language Processing (REFLEX)



Motivation for REFLEX

- ◆ Analysts and automated analytic tools need access to information conveyed by natural language in documents
- ◆ Convergence of programs and requirements
 - ITIC Machine Translation research initiative, includes focus on Essential Elements of Information
 - Automatic Content Extraction (ACE) program, Phase 3
 - Requirement to assist analysts in navigating through heterogeneous natural language data sets (various languages, media, genres, domains)
 - Requirement to feed structured data objects to automated analytic tools, based on information in natural language documents



Program Goals

- ◆ Assistance to Information Analysts in:
 - Judging relevance quickly for text or ASR/OCR output
 - Judging relevance from foreign language
 - Knowledge discovery by clustering documents having certain elements in common
 - Identifying specific facts of interest in text
 - (longer term) performing information analysis from translation



Program Goals (cont'd)

- ◆ Input to automated analytic tools
 - Convert information in a document or document cluster to a language-independent knowledge representation or data structure
 - Support SNA, Knowledge Discovery, KB population
 - Separate language processing issues from analytic tools

ACE Tasks - Relation Detection - Focused

Example News Broadcast Text:

“On Saturday night, **the Glass House dance club** in Pomona was packed with 800 people enjoying the holidays when the **local police** and members of a Los Angeles County **hazardous materials team** walked in and stopped the music ... A few minutes earlier someone had called **911** and reported what has become a common threat ... The caller said **the lethal bacteria anthrax** had been released in **the club.**”

Example Extracted Information:

```
<AT>
  What: <FACILITY ID=F1>
  Where: <GPE ID=G1>
</AT>
```

```
<AT>
  What: <PERSON ID=P2>
  Where: <FACILITY ID=F1>
  When: “Saturday night”
</AT>
```

```
<PART_OF>
  Parent: <GPE ID=G2>
  Child: <ORG ID=O2>
</PART_OF>
```

```
<AT>
  What: <PERSON ID=P1>
  Where: <FACILITY ID=F1>
  When: “Saturday night”
</AT>
```

```
<ROLE>
  Pers: <PERSON ID=P2>
  Org: <ORG ID=O2>
  Role: “member”
</ROLE>
```

```
<PART_OF>
  Parent: <GPE ID=G2>
  Child: <ORG ID=O3>
</PART_OF>
```

```
<AT>
  What: <ORG ID=O1>
  Where: <FACILITY ID=F1>
  When: “Saturday night”
</AT>
```

```
<AT>
  What: <SUB ID=S1>
  Where: <FACILITY ID=F1>
  When: “Saturday night”
</AT>
```

Analytic browsing example

The Thai ambassador to the UN warned that Indian food is fattening

As the Cricket World Cup drags on into the semi-finals, the 5-day test match between New Guinea and Scotland is showing no signs of ending

The announcement by the National Ketchup Council that powder-milk biscuits are low in anti-oxidants has ignited street protests in Lake Woebegon.

The president of Upper Slobovia believes that the end of the world is near and is making preparations as needed.

John Allen Smith
Fred Armbruster
Timbuktu
Feb 29th, 2004
Jose

Vehicle
Carburator
End-of-year Sales
Gross revenue

Upper Slobovia is in an existential crisis

President Slob of Upper Slobovia has stated that the Armageddon is nigh

President Slob Upper Slobovia Armageddon Slobville

UN Observers have not detected that the sky iIs falling anywhere in Upper Slobovia

Western editorials have not credited the Upper Slobovian the possibility that he may be right

The Upper Slobovian Ministry of Information Reported yesterday that their beloved leader, President Vincent Slob, has determined that the end of the world is fast approaching. Based on his years of cabalistic studies, numerology, Bayesian and Demster-Schaefer Reasoning, and phrenology, he has calculated the odds of the end of the world occurring in the next three weeks as 23%

Περ ασπερα αδ
Αστρα, διχιτ ιλ
Φαβυλα δε Πετρο
Χυνιχυλο. Χοντρα
Σπεμ Σπερο, Δει
Συβ νυμινε πιγετ,
Ετ Ε Πλυριβυσ Υνουμ.
Δε ρεσ νον αδ ρορυμ.
Ιλλεγιτιμι νον
Χαρβορυνδυμ.
Νον ινηιβιτο
Ρατιονυμ.



Research Areas

- ◆ In the BAA & PIP, tasks are not grouped according to technology or capability (such as MT, Extraction, etc.)
- ◆ The BAA and PIP group tasks in terms of semantic coverage
 1. Partial semantic content: Entities
 2. Partial semantic content: Relations
 3. Partial semantic content: Events
 4. “Full” semantic content
 5. Linguistic resource creation and validation
 6. Evaluation methodology research
- ◆ The “Research Areas” part of the PIP identifies specific subtasks for these 6 areas.
- ◆ Proposal must identify the sections, subsections, and subsubsections (from the PIP) that they are addressing



Data

◆ Languages

- English, Arabic, Chinese, (probably) Korean
- Less-Commonly-Taught Languages (LCTLs)

◆ Media

- Text, and derived text / intermediate output from OCR and ASR

◆ Domains and Genres

- No domain limitation
- Newswire/newspaper reporting, technical journal articles/abstracts, and broadcast news / conversational speech transcripts (automatic or ground-truth)

◆ Applications

- Broad range of possible down-stream applications
- Social Network Analysis is a good example of one such application; SNA is considered in the selection of semantic types to be handled



What REFLEX is not:

- ◆ It is NOT an ASR research program
- ◆ It is NOT an OCR research program
- ◆ It is NOT a SNA research program
- ◆ It is NOT a commercial product development program



QUESTION #1

Q: Is the program open to non-US companies?

A: Proposals will be entertained from any R&D organization located in the US. These organizations, however, may choose to propose subcontracts with non-US entities



QUESTION #2

Q: Can new start-ups submit proposals?

A: Yes



QUESTION #3

Q: Can we submit proposals for language X or a language from country Y?

A: The main program languages are English, Chinese, Arabic, Korean.

Most of the tasks for LCTLs address developing algorithms or methodologies for handling arbitrary LCTLs, not for working with a specific LCTL. Only the tasks in Sections 4.1, 5.2.1, 5.2.2, and 5.2.3 allow the bidder to propose work on a specific LCTL. There is no restriction in the BAA or PIP on which LCTLs can be proposed under those tasks, however.



QUESTION #4

Q: When is the Industry Day / Question and Answer session for this BAA?

A: Now



QUESTION #5

Q: Is the BAA restricted to pure research, or is there interest in applied research leading to an actual working system?

A: We are interested in applied research leading to prototypes or demonstration systems, but do not expect to fund productization of commercial products. Volume 1, Part II, Section f of your proposal should address Deliverables



QUESTION #6

Q: Would it be appropriate to propose topic XYZ under this BAA?

A: Sorry, you have to decide. Sections 1 through 6 of the “Research Areas” in the PIP identify allowable topics.



QUESTION #7

Q: Can non-US citizens be on a proposal?

A: The BAA places no restriction on the citizenship of key personnel



QUESTION #8

Q: Can proposals be submitted via e-mail or a web page?

A: No



QUESTION #9

Q: When are proposals due?

A: April 30th, 2004, 4:00 pm Mountain
Standard Time



QUESTION #10

Q: The BAA states: “This BAA will remain open for 3 years after the publishing date.” What does this mean?

A: Funding has been made available for Research on English and Foreign Language Exploitation. An initial round of proposals is expected by 30 April 2004, and these funds are earmarked for this first round. However, when, and if additional funds become available in the future, late proposals could be funded at that time. After 12 March 2007, no more proposals can be considered under this BAA, even with speculative funding. Any proposal received between 4:01 p.m. on April 30, 2004 and March 12, 2007 is in this category of "late".



QUESTION #11

Q: Can graduate students attend the Program meetings? Can we propose conference travel for graduate student Research Assistants?

A: We strongly encourage PI's at universities to propose travel for relevant graduate student RA's to Program meetings, evaluation conferences, etc. Propose travel costs appropriately.



QUESTION #12

- ◆ Q: What do you mean by “Teaming between industry, independent research centers and academia is encouraged”? Do we have to team?
- ◆ A: No, teaming is certainly not required. If it makes sense for your proposed work to put together a team, then team. Teams of multiple academic organizations or multiple industry labs are also completely acceptable.



QUESTION #13

Q: What is ITIC?

A: The Intelligence Technology Innovation Center sponsors research in many fields for the Intelligence Community



QUESTION #14

- ◆ Q: How does the Government intend to evaluate the performance and effectiveness of the new technology to be developed?
- ◆ A: Because of the range of tasks covered under REFLEX, the Government will not necessarily offer community-wide evaluation for each task or technology. Certain tasks will continue to be addressed via existing evaluation programs, such as the ACE evaluation and NIST's MT Evaluation. Certain other task-specific evaluations will emerge over the course of the program (partly in response to the balance of proposals that are funded). However, each proposal must discuss the offeror's plans for internal evaluations of progress, in proposal Volume 1, Part II, (a), Technical Approach. The Government expects to pursue a separately-funded effort to integrate component technologies from various Program participants to demonstrate certain capabilities.



Question #15

- ◆ Q: When will the LCTLs be decided?
- ◆ A: The exact LCTLs to be worked will be decided in consultation with the successful bidder(s) after award of contract. The language for L4 will be finalized toward the end of the first year of performance.



Question #16

- ◆ Q: Can we hear more about who is the customer for the REFLEX BAA (Agency name, Agency's general mission and/or more specific application, etc.)
- ◆ A: REFLEX addresses the needs of a number of Intelligence Community, Defense, and Homeland Security agencies.



Question #17

- ◆ Q: You mention that FFRDC's are welcome. Would you prefer for them to be on a Vendor team, Lead the proposal or participate on a particular part of the response?
- ◆ A: No preference.



Question #18

- ◆ Q: Do you expect the respondents to put equal weight into A) Development and B) Evaluation?
- ◆ A: This is a business decision on the part of the proposer. Balance your efforts as you feel is appropriate.



Question #19

- ◆ Q: Would you rank by priority the priorities of the BAA requirements and/or corresponding technology research area?
- ◆ A: All of equal weight.



Question #20

- ◆ Q: How much funding is available for 2004? How much funding is available for ongoing years? How many awards do you expect for this BAA per year and/or on an ongoing basis? what proportions of the total funding will be used for multiple awards each year and/or over several years?
- ◆ A: We expect to make 10-15 awards in FY04. The Government may not necessarily exercise third year options on all funded efforts. Additional awards in future years are possible as additional funding becomes available.