

Developing and Evaluating a Document Visualization System for Information Management

Bowen Hui

Dept of Computer Science

University of Toronto



Motivation for Modeling and Filtering Information

- much of human knowledge is expressed in free text
- impose “document model”
- conceptual modeling (CM) concerns with E-R
- push CM towards semi and unstructured data
- issues:
 - level of representation
 - complexity of processing technique
- we propose simple CM that allows efficient extraction
- evaluate new CM’s usefulness



Motivation for New Evaluation

- current accuracy metrics:
 - no single correct answer
 - misleading/uninformative for the untrained consumer or avg user
- current evaluation methods:
 - time consuming and labour intensive
 - requires translation/linguistic experts
 - cognitively demanding tasks
 - too often influenced by evaluator's intelligence
 - inconsistent (unreliable) results
 - does not extend across systems or domains

Want direct feedback on user acceptance of output quality and areas of focus for researchers or developers.

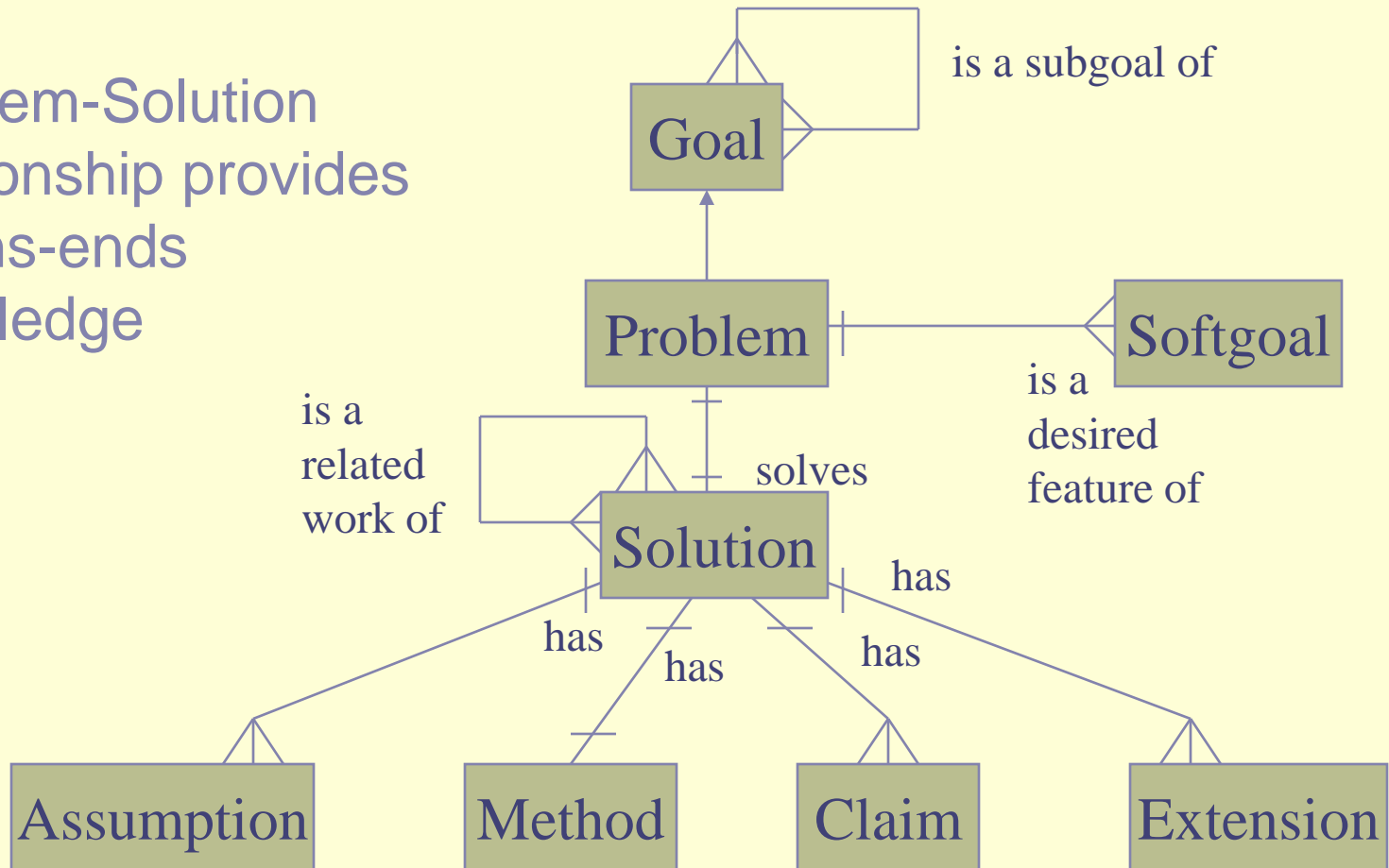


Outline

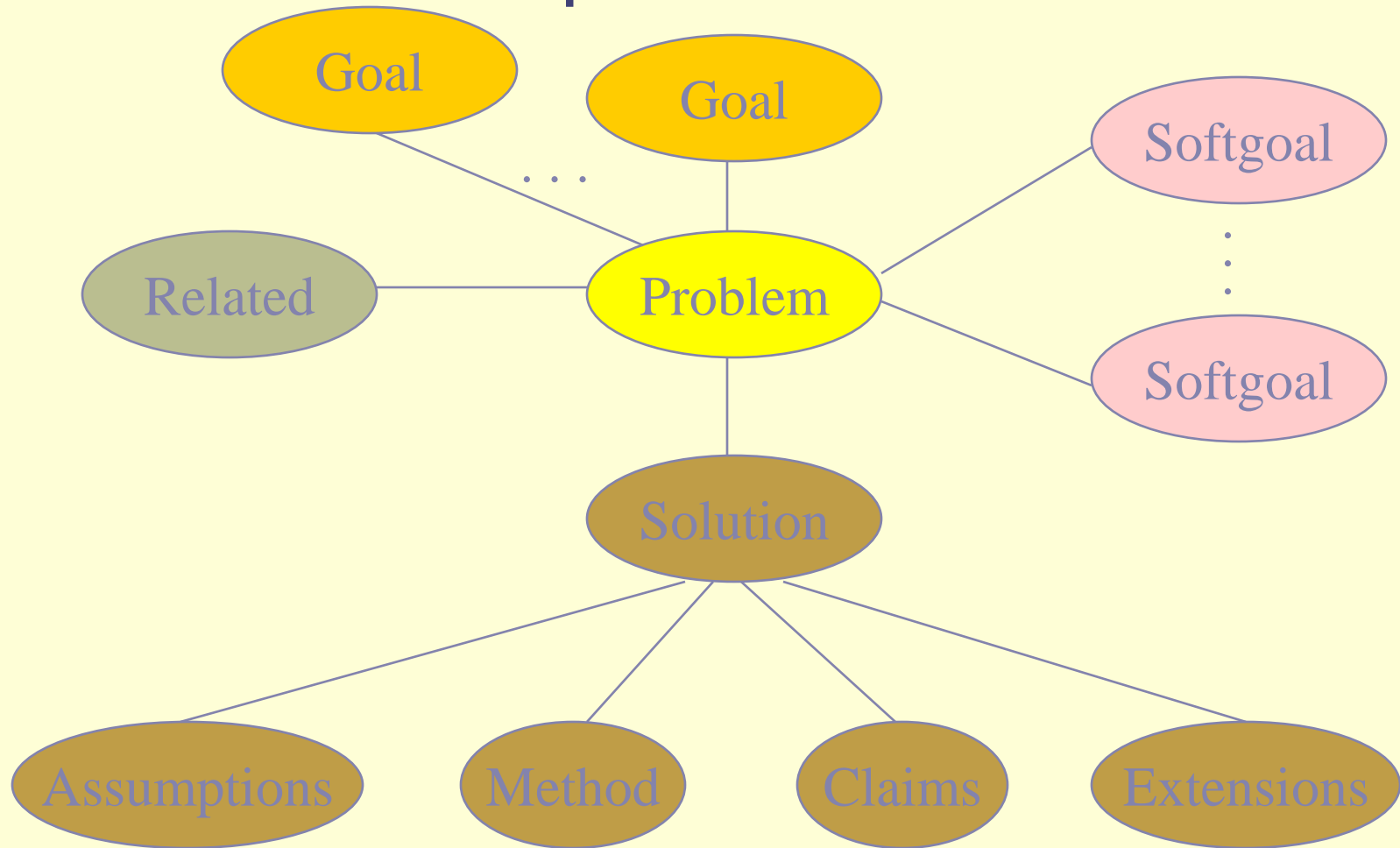
- proposed conceptual model for specialized documents
 - patents
 - scientific articles
- system implementation
- empirical evaluation
- summary of contributions

Document Conceptual Model

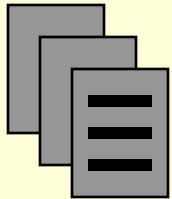
- Problem-Solution relationship provides means-ends knowledge



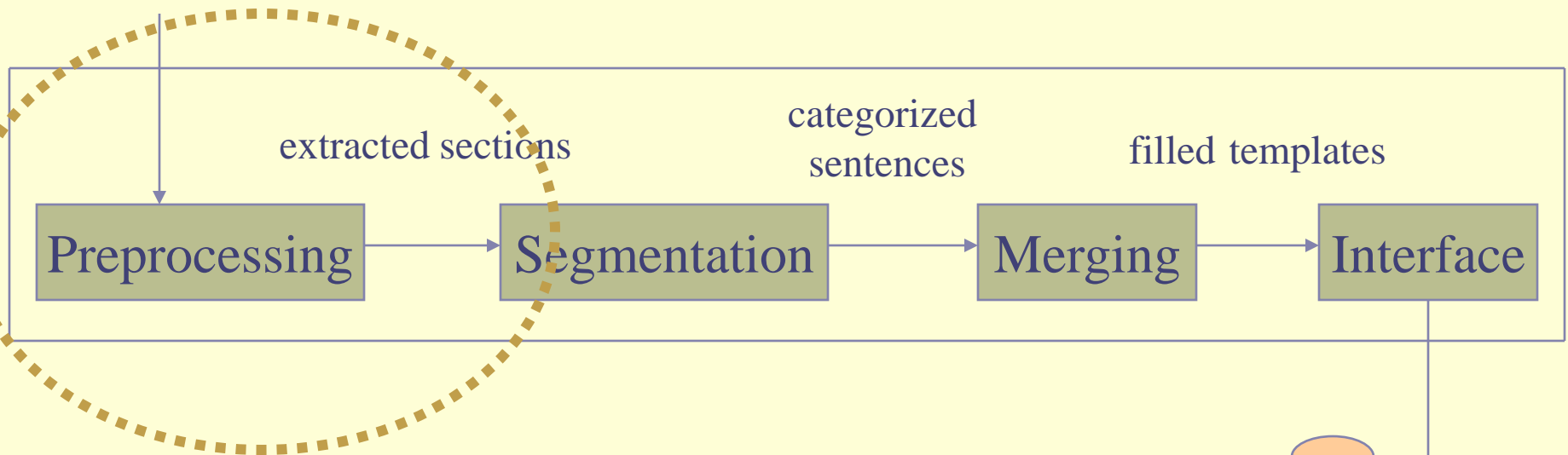
Graphical Model



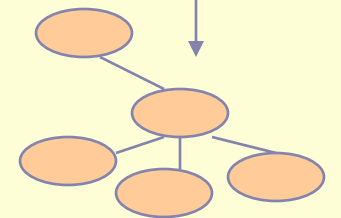
System Components



raw texts

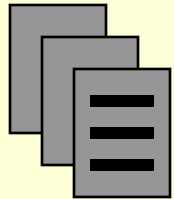


- pat: abstract + summary
- sci: abstract, introduction, conclusion



graphical layout

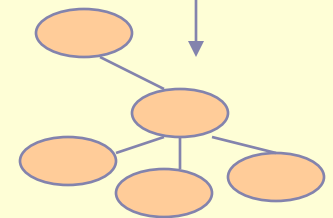
System Components



raw texts



- categorize sentences
- inherit if no match
- categorize phrases (GOAL and SOFTGOAL)
- assign OTHER if unmatched



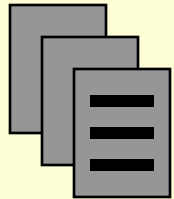
graphical layout

Template Slots and Patterns

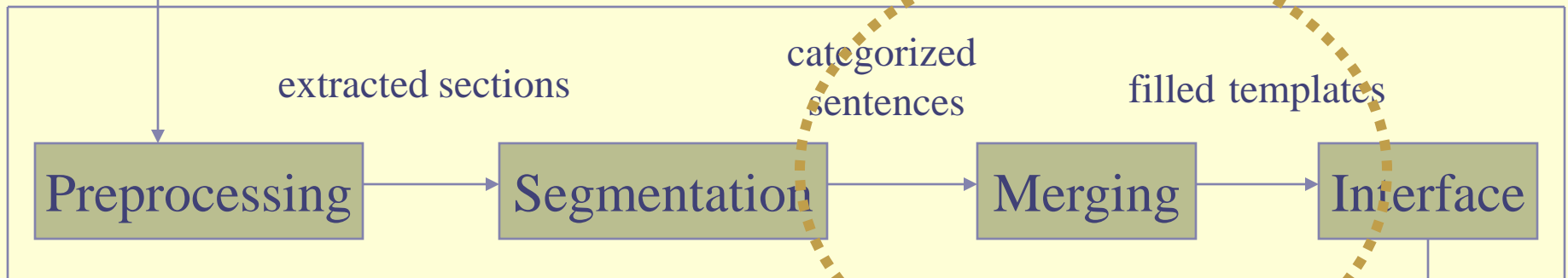
PROBLEM	“Our objective in ... is the study of”
RELATED	“Another successful approach”
SOLUTION	“We propose”
ASSUMPTIONS	“A long standing objection”
METHOD	“to achieve ... is used”
CLAIMS	“significant improvement”
EXTENSIONS	“yet to be investigated”
OTHER	“In this section, we”
GOAL	“technology in education”
SOFTGOAL	“immediate feedback”

● used 20 documents to develop 316 rules (Perl regex)

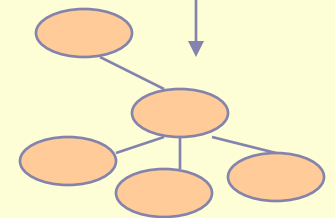
System Components



raw texts



- for each slot,
 - remove common/transition phrases
- for each stemmed sentence pair,
 - delete if similar

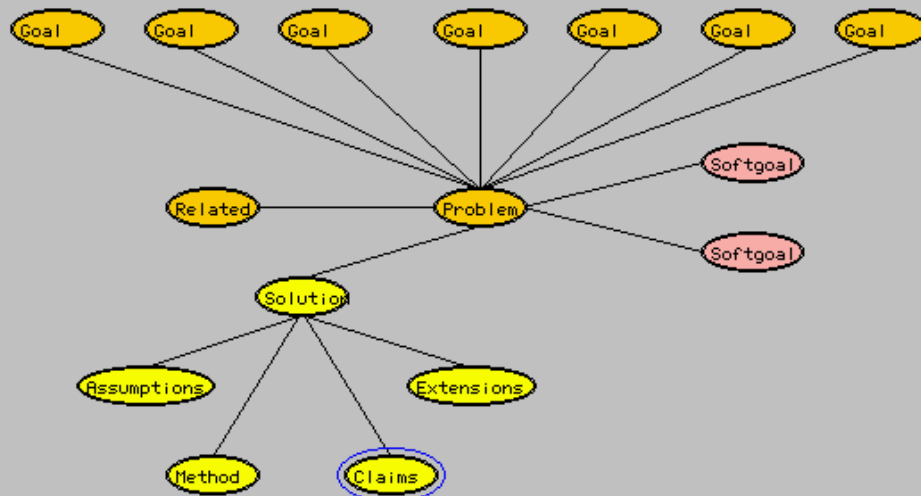


graphical layout

System Interface

Title: Children as Our Technology Design Partners

Graphical View



Claims and Results:

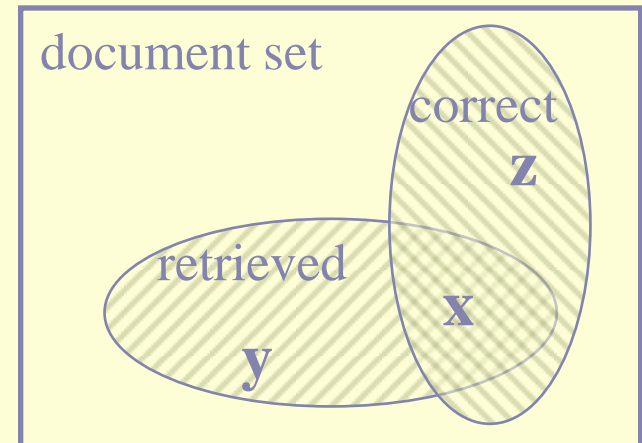
- we need to continue to refine the research process with children for children, because ultimately our goal is simple: to create exciting, meaningful new technologies for children.
- we need to establish new development methodologies that enable us to stop and listen, and learn to collaborate with children of all ages.
- in the chapter that follows, a discussion of new research methodologies will be presented.
- we believe that this has been the case because school activities lend themselves to the existing observation and participation methodologies.
- schools are generally places where children are asked to carry out directed, a dult-specified tasks.
- children are typically not in control of when they can have art or what they can write about, even when they can go home. Ultimately, we believe that researchers can only tell so much about what children want or need in technologies from environments such as these.
- we are trying to understand how we can bring our knowledge from the “real world” of children into the “design world” of technology development.

Empirical Study

- evaluate our conceptual model
- accuracy of extraction (segmentation)
- intercoder reliability
- final coder evaluated each segmentation output
- full eval on 93 new docs

$$\text{recall} = \frac{x}{y \text{ (seg. categories)}}$$

$$\text{precision} = \frac{x}{z \text{ (coder categories)}}$$



Accuracy: Results

Domain	# Docs	# Sent	Precision	Recall
colour toy	11	324	73.98	72.84
education mathematics	12	430	81.35	81.16
design blouse	12	400	65.80	63.50
modern chair	12	311	67.18	70.42
Patents	47	366	72.08	71.98
women and language	10	527	37.83	38.33
children HCI	12	791	55.66	55.66
interface HCI	12	502	51.28	51.79
computational linguistics	12	640	49.37	49.06
Scientific Articles	46	615	48.53	48.71



Adopting Usability Evaluation

- **Heuristic evaluation** (Nielsen 1993)
- inspection method that walks through system by carrying out meaningful task in real-use scenario
- familiarize evaluator w/system capabilities
- evaluation centred around a set of principles
- resulting comments = list of identified sys probs
- benefits:
 - easy set-up
 - low time demand
 - few evaluators (~5 to identify 75% system errors)



Adapting to NL Evaluation

- cast filtering tool as basic text summarizer (TS)
- evaluation of system *output* quality, not interface
- in addition to commenting on principles, also want quantitative *score*
- for TS domain, we need:
 - principles
 - meaningful task
 - test materials
- more thorough test-bed in MT domain



TS: Principles

1. **Conciseness:** not irrelevant, redundant
2. **Retention:** as compared to original document
3. **Coherence:** within sentence, overall summary
4. **Consistency:** words/concepts
5. **Informativeness:** useful, easily accessible
6. **Comprehensibility:** easy to understand
7. **Fit for Audience:** experts/novices
8. **Fit for Purpose:** fits intended task and purpose

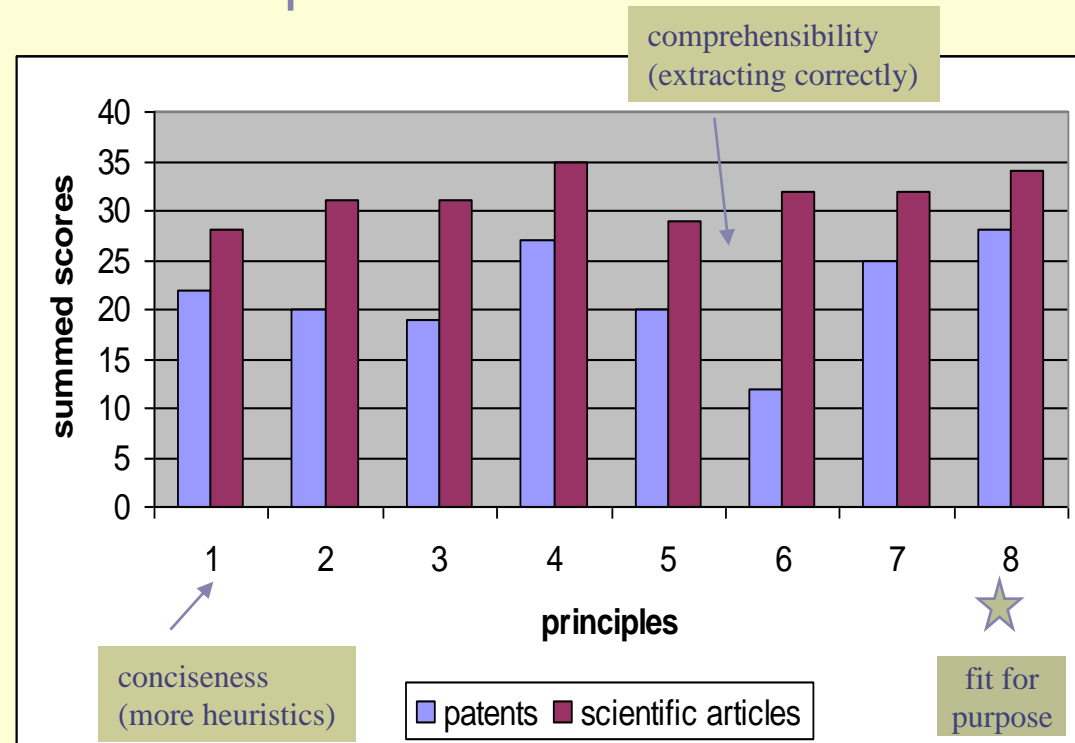


TS: Task

- models a reviewer's question-answering task
 1. What is the problem addressed by this work? Does it describe why the problem is significant?
 2. Does the work present the approach taken to solve the problem targeted? Is the design or implementation of a system described in terms of key ideas of the approach?
 3. What are the contributions of this work? Are the benefits and limitations clear? Are the results positive or negative?

TS: Experiment and Results

- pilot (3 participants) evaluated P1,S1 (not shown)
- full (4 participants) evaluated P2,P3,S2,S3
- complete task for summarized outputs
- for each output, rate each principle out of 5
- comments, access to original and summarized texts
- time spent between 1 hour – 2 hour

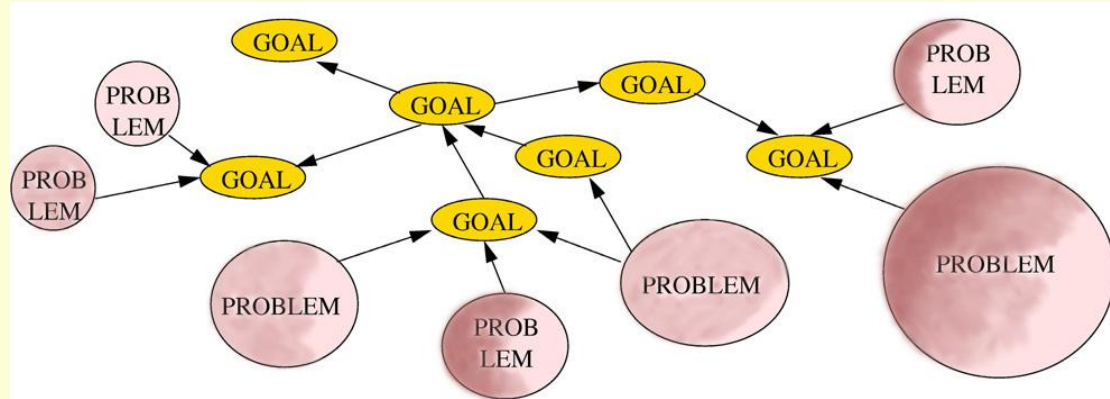




Summary: Modeling Work

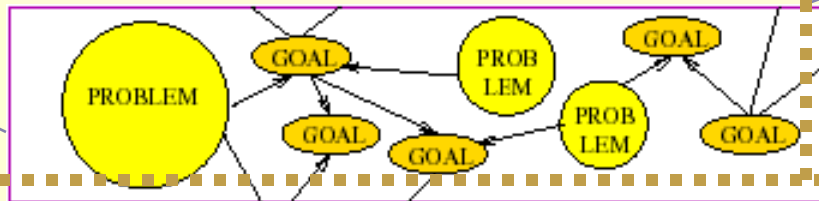
- we motivate a simple CM to model 2 kinds of free text documents
- concepts based on user-centred goals
 - relations: problem-solution, benefits and drawbacks, assumptions, methods, claims, extensions
- capture high-level argumentative structure
- evaluated CM *within system implementation*
- generalize over document types by grouping patterns that suit the writing styles of each
- users get value from simple modeling and processing of our approach

Visionary System



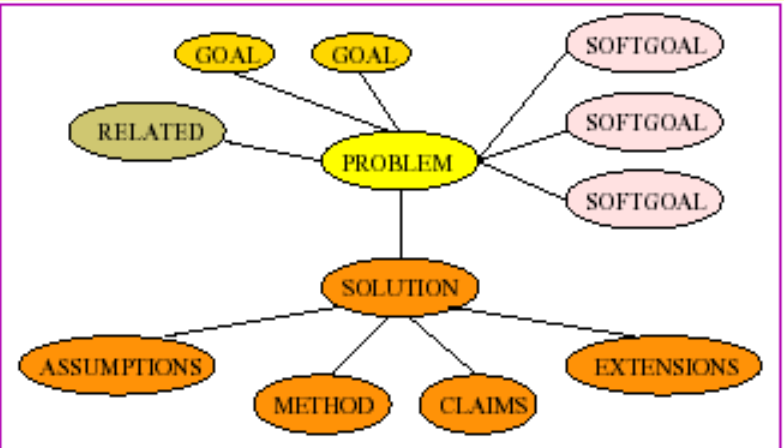
network of goal and problem nodes

visionary system interface



EXTENSIONS

- it is dangerous, but necessary, to dream about the future
- danger because misguided dreams mislead designers, necessary because without vision navigation is difficult
- without dreams we risk stagnation, and lose the chance to make a better world
- there is excitement in the user interface research and development community ideas are emerging dialy and the reduction to practice is rapid





Summary: Evaluation Work

- need for measuring usability and getting at development problems of NL systems
- adapted heuristic evaluation
 - 1. preliminary test-bed in TS
 - 2. comparative MT evaluations
- major experienced advantages:
 - not cognitively overwhelming for evaluators
 - easy evolution of principles and task according to application

Heuristic evaluation works well for NL tasks
without a gold standard



Principled-based Customization

- establish summarization/translation principles
 - through usage and standardization
- component selection
 - combining the “best” components of different systems
- user profiling
 - collect data where users rank the evaluation principles to reflect which criteria are more important to them
 - this also gives an indication of errors that are more *forgiving*
 - create user groups based on criteria (vs. demographics)