

Conversation for Textual Case-Based Reasoning

Kalyan Moy Gupta

KNEXUS
RESEARCH CORPORATION

David W. Aha



Naval Research Laboratory

Workshop on Textual Case-Based Reasoning: Beyond Retrieval,
ICCBR-07, Belfast, Northern Ireland, 15 August 2007

Copyright© Knexus Research Corp., 2007

Introduction

Introduction

Problem

TCB Engg.

Conversation

Solution

Conclusion

Textual CBR

Method of reasoning with cases derived from artifacts containing blocks of *unrestricted natural language text*

Artifacts

Incident Reports

Legal Briefs

Emails

Textual Case-Base Engg.

- 1.Feature Extraction
- 2.Feature Vocabulary Development
- 3.Feature Assignment/Indexing

Textual Case Base

Indexing Vocabulary (Terminology)

Problem

- ✓ Introduction
- ▶ Problem
- TCB Engg.
- Conversation
- Solution
- Conclusion

- Many TCBR applications such as industrial troubleshooting and product recommendation require conversation
- Existing methods for conversation only utilize hand-crafted features (i.e., manually engineered cases). However, practical and cost effective TCBR systems should conduct conversation with automatically generated features
- No suitable methods are currently available

Our Goal

1. **Define a new research problem** of conversation for TCBR systems using automatically engineered textual cases
2. **Propose a new methodological framework** for designing TCBR conversations

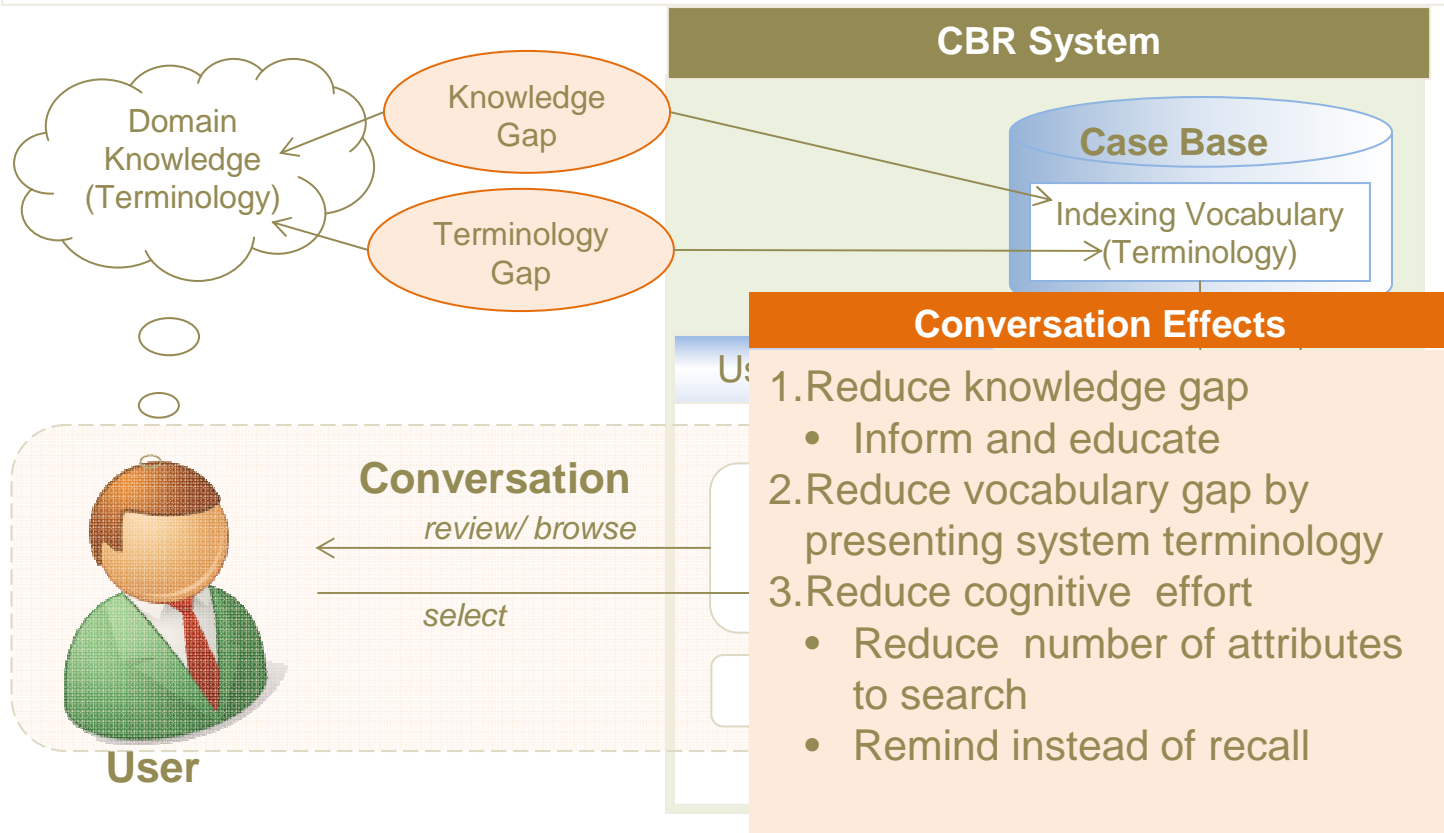
Textual Case Base Engineering (TCBE)

	TCBE Task	Manual	Automatic
✓ Introduction	Extract Attributes Terms and phrases (e.g., Cessna, Seaplane)	Identify key phrases based on <i>domain knowledge</i>	<ul style="list-style-type: none"> String tokenization Phrasal parsing and distribution analysis
✓ Problem			
▶ TCB Engg.	Identify Values (e.g., Aircraft has values {Cessna, ..., Seaplane})	Use <i>domain knowledge</i> to identify attribute value relationships	Unavailable. Could use ontologies (e.g., WordNet)
Conversation	Select Attributes (Reduce 1000s of attributes to 10s or 100s)	Observations and Conjectures <ol style="list-style-type: none"> Manual methods produce a more compact, richly organized, and more lucid vocabulary than automatic indexing due to extensive use of domain knowledge and attribute synthesis Manually generated vocabulary is likely to be superior to automatic methods, especially when the case base is derived from relatively few unlabelled verbose documents (e.g., incident reports) Manual method is impractical except for simple, relatively static domains 	
Solution	Organize Attributes (e.g., Taxonomies and semantic networks)		
Conclusion	Construct Attributes (Introduce a new term or concept)		
	Index Cases (Assign features to cases)		

Conversation

- ✓ Introduction
- ✓ Problem
- ✓ TCB Engg.
- ▶ Conversation
- Solution
- Conclusion

Cooperative exchange of attribute suggestions (prompts) and value selections between users and CBR systems to formulate effective queries



Conversation: Performance Measures

- ✓ Introduction
- ✓ Problem
- ✓ TCB Engg.
- ▶ Conversation
- Solution
- Conclusion

Effectiveness

Retrieval performance of a query formulated by conversation compared to the retrieval performance of the best query (e.g., problem representation of cases with target solution)

$$\text{Effectiveness} = \frac{f_r(\text{CB}, q_c) : \text{Conversational Query Retrieval Performance}}{f_r(\text{CB}, q_b) : \text{Best Query Retrieval Performance}}$$

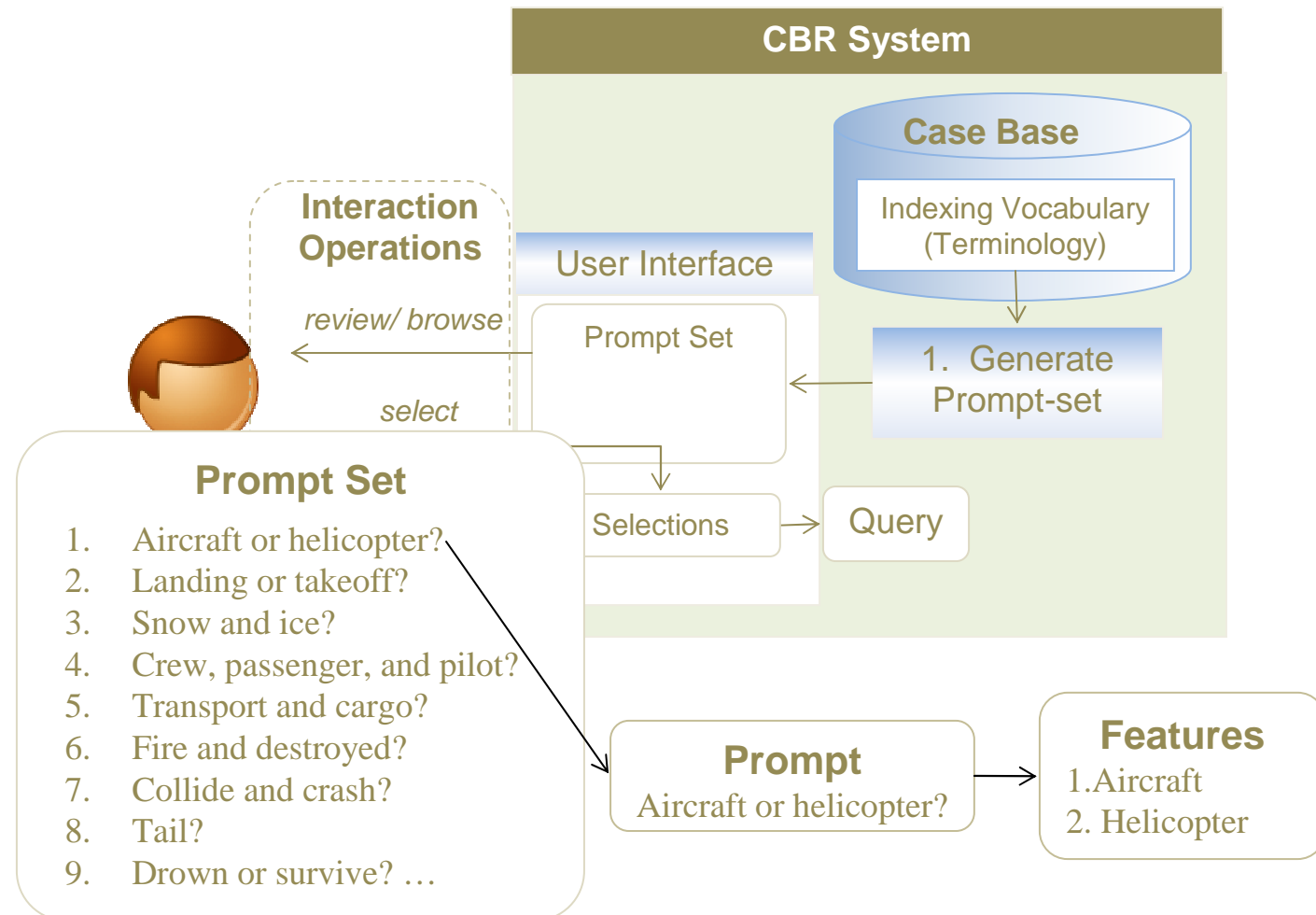
Efficiency

The number of browse and select operations performed per feature in a query

$$\text{Efficiency} = \frac{\# (\text{Select operations}) + \# (\text{Prompts})}{\#(\text{Features in a query})}$$

TCBR Conversation

- ✓ Introduction
- ✓ Problem
- ✓ TCB Engg.
- ✓ Conversation
- ▶ Solution
- Conclusion



TCBR Conversation: Prompt Types

	Prompt Type	No of features	Impact on conversational performance and user interaction		
			Effectiveness	Efficiency	Flexibility
✓ Introduction					
✓ Problem	One-of-one: Allows selection of one independent feature	1	lowest	low	highest
✓ TCB Engg.	<ul style="list-style-type: none"> •Rudder? •Aerodrome? 				
✓ Conversation	One-of-many (XOR): Allows selection of one feature from a set of multiple mutually exclusive features	n	high	moderate	moderate
▶ Solution					
Conclusion	All-of-many (AND): Allows selection of a set of n <i>strongly</i> correlated features	n	highest	highest	lowest
	<ul style="list-style-type: none"> •Water, submerge, and drown? 				
	Some-of-many: Allows selection of m-of-n features from a set of moderately correlated features.	n	moderate	high	moderate
	<ul style="list-style-type: none"> •Fire? [] Destroyed? [] Damaged? [] 				

TCBR Conversation: Prompt Generation

	Generation Subtask	Methods (Dynamic and/or Cached Modes)	Performance Impact	
			Effectiveness	Efficiency
✓ Introduction	Feature Selection	<ul style="list-style-type: none"> Unsupervised: Wiratunga <i>et al.</i>, ECCBR-06, and TF/IDF ... Supervised: Info-Gain, Rough Sets, ... 	Highest	Highest
✓ Problem				
✓ TCB Engg.	Feature Interrelationship Identification	<ul style="list-style-type: none"> XOR: General and domain ontologies AND: Mutual information, etc. Correlation (Weak conjunction): Mutual information, and association rule mining techniques 	Moderate	Moderate
✓ Conversation				
▶ Solution	Prompt Composition	<ul style="list-style-type: none"> Using templates and feature interrelationships 	?	None
Conclusion	Prompt Ranking	<ul style="list-style-type: none"> Multi-attribute prompt scoring and ranking 	?	?

Conclusion

- ✓ Introduction
- ✓ Problem
- ✓ TCB Engg.
- ✓ Conversation
- ✓ Solution
- ▶ Conclusion

Contributions

1. **New research problem definition** for conversational TCBR systems using automatically generated feature vocabularies
2. **New methodological framework** for designing effective and efficient TCBR conversations
 - Performance measures, prompt types, and their impact on conversational performance, and prompt generation tasks and associated methods.

Future Work

Evaluate proposed framework by developing and implementing selected combinations of prompt generation methodologies

Discussion

- ✓ Introduction
- ✓ Problem
- ✓ TCB Engg.
- ✓ Conversation
- ✓ Solution
- ✓ Conclusion

Thank You!

Conversation for Textual Case-Based Reasoning

Kalyan Moy Gupta

KNEXUS
RESEARCH CORPORATION

David W. Aha



Naval Research Laboratory