

Indexing strategies for goal specific retrieval of cases

Michael J. Pazzani
Department of Information and Computer Science
University of California,
Irvine, CA 92717
pazzani@ics.uci.edu

One of the most curious features of the history of economic sanctions has been the the extent to which the experience of past cases has been overlooked or ignored-- Robin Renwick

Outline

- I. Problem: retrieval of individual cases for multiple purposes
 - A. Explanation
 - B. Prediction

- II. Why indexing is the right way to think about retrieval
 - A. Psychological-
 1. Reconstructive memory
 2. Encoding specificity & levels of processing
 - B. Computational Retrieving relevant rather than similar cases

- III. Indexing strategies for goal specific retrieval of cases
 - A. Retrieval of individual and generalized cases
 1. Type of processing done at storage time
 2. Content of retrieval cue
 3. Goal of retrieval
 - B. Indices
 1. Surface (rather than abstract) features
 2. Strategic goal (explain vs. predict)
 3. Feature to be explained or predicted

Question: What would happen if the United States refused to sell computers to South Korea unless South Korea stopped exporting automobiles to Canada?

OCCAM¹: The goal of the United States that South Korea not sell automobiles to Canada will fail and South Korea will purchase computers from a country which exports computers. This happened when the United States did not sell grain to the Soviet Union after the Soviet Union invaded Afghanistan. Argentina sold grain to the Soviet Union. Also, Australia did not sell uranium to France after France exploded nuclear weapons in the South Pacific. South Africa sold uranium to France.

1. OCCAMS output is edited slightly, by adding tense information to verbs.

Question: What would the United States do if Turkey invaded Cyprus?

OCCAM: The United States would refuse to sell a commodity to Turkey if Turkey invaded Cyprus. This happened when Greece invaded Bulgaria. The League of Nations refused to sell food to Greece. Also, when the Soviet Union invaded Afghanistan, the United States refused to sell grain to the Soviet Union.

Why indexing is the right way to think about retrieval:

Reconstructive memory

(Bartlett, 1932; Schank 1980; Kolodner 1984)

- A case is organized in memory by the knowledge structures that guided the comprehension of case.

Corollary: A case will be indexed in multiple places if more than one knowledge structure is accessed during comprehension

- Only those aspects of a case which differ from those of the knowledge structure used for comprehension are stored.

Differing features become indices that allow retrieval of the case when that feature is present in the retrieval cue (or inferred from information in the retrieval cue)

- As more cases are added to memory, additional knowledge structures are created to organize new experiences.

Why indexing is the right way to think about retrieval: Experimental Evidence

- Semantic memory (Freedman & Loftus, 1971)
Category (e.g. vegetable) & attribute (e.g., green)

Subjects presented with category before attribute retrieve exemplar more rapidly than attribute before category.

- Episodic Memory (Reiser, Black & Abelson, 1987)
Activity (e.g. Eating at a Restaurant) &
action (didn't get what you want)

Subjects presented with activity before action retrieve exemplar more rapidly than action before activity.

The effect of processing task

- Levels of processing (Craik & Lockhart, 1972)
Retrievability affected by depth of processing during storage
Jacoby & Dallas (1981):
Subjects study word list: Subjects answer question
 1. Appearance: Is it in capital letters
 2. Sound: Does it rhyme with "teach"
 3. Meaning: Is it a form of communicationIn test phase: subjects verify that they studied word.
Success Rate: Meaning 50% < Sound 63% < Meaning 86%
- Encoding Specificity (Tulving, 1983)
Retrieval affected by similarity between storage and retrieval context

Recognition failure of recallable words.

Seifert (1988) found priming of story recall related to processing goal at storage and retrieval

Why indexing is the right way to implement CBR:
Retrieve the most relevant rather than the most similar case

- An index is a feature of a case identified at storage time that describes the situations in which the case should be retrieved.
- Proposals by Waltz and Thagard & Holyoak cannot account for situations in which there is a unique item specified by the retrieval cue in memory, but it is not retrieved.

“Can you think of an incident in which the US threatened a third-world country and a US adversary helped out?”

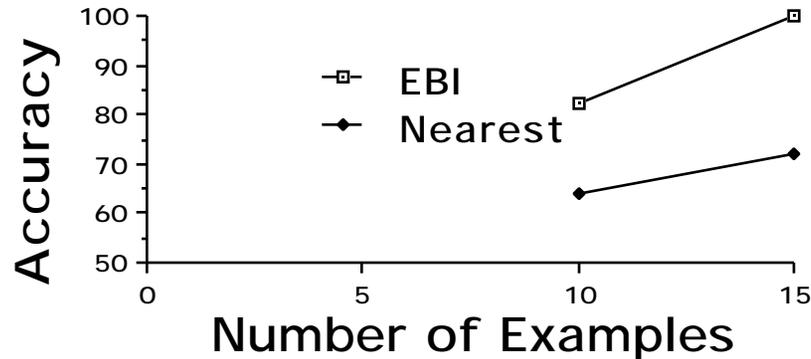
“Can you think of an incident in which an English speaking country threatened an African country and a country that exports palladium helped out?”

- Proposal by Waltz assumes that the memory processes cannot distinguish most similar case (according to the surface features) from the most relevant.

Experimental Comparison of nearest neighbor vs. EBL

Input: Description of 10 or 15 economic sanctions incidents.

Test: Predict success/failure of five hypothetical incidents.



1961 USSR vs. Albania

threat: Refuse to sell grain

demand: Stop economic ties with China

outcome: (fail) China sells Albania

Canadian wheat at lower price

1976 US vs. Ethiopia

threat: Stop aid (57 million)

demand: Stop human rights violations

outcome: (fail) Soviets provide aid

1980 US vs. USSR

threat: Cut off grain sales

demand: Withdraw troops from
Afghanistan

outcome: (failure) Buy grain from

Argentina at a higher price

1983 Australia vs. France

threat: Not sell uranium

demand: Stop nuclear tests

outcome: (fail) France buys from

South Africa at higher price

What makes a good index?

I. Relevant surface features

A. Hundred step rule-

Deep features (e.g., morals) cannot be computed in time

B. "Morals and effects are the outputs of the retrieval process, not its inputs"

C. Relevant surface features are those surface features that allow the solution to a problem to be inferred.

II. A feature relevant for one task may not be relevant for a different task. Processing task at storage time becomes part of index

A. Explanatory Indices

"What could cause the price of oil to rise?"

B. Predictive Indices

"What would happen if the US refused to sell computers to South Korea if South Korea continues to export automobiles to Canada?"

What makes a good index?

III. Different features of the same case can be explained or predicted. The feature that needs to be explained or predicted becomes part of the index.

A. Predicting the result of an action

What would happen if the United States refused to sell computers to South Korea unless South Korea stopped exporting automobiles to Canada?

B. Predicting what action will occur

What would the United States do if Turkey invaded Cyprus?

Summary: An index consists of a triple

- feature and value of a case (or generalized case)
- description of task_{explain or :predict}
- feature to be explained or predicted.

Deriving Indices

I. Feature and value (What features are predictive)

- A. Analytically: EBL
- B. Empirically: SBL

II. Task

- A. Analytically: antecedent/consequent of domain knowledge
- B. Empirically: predictiveness vs predictability

III. Feature to be predicted (Or explained)

Determined by processing goal

Determining Indices for US-USSR grain embargo:

Processing goal: Explain result (USSR buys grain from Argentina)

1. Threat -> Increased demand
2. Increased demand -> Willingness to pay higher price

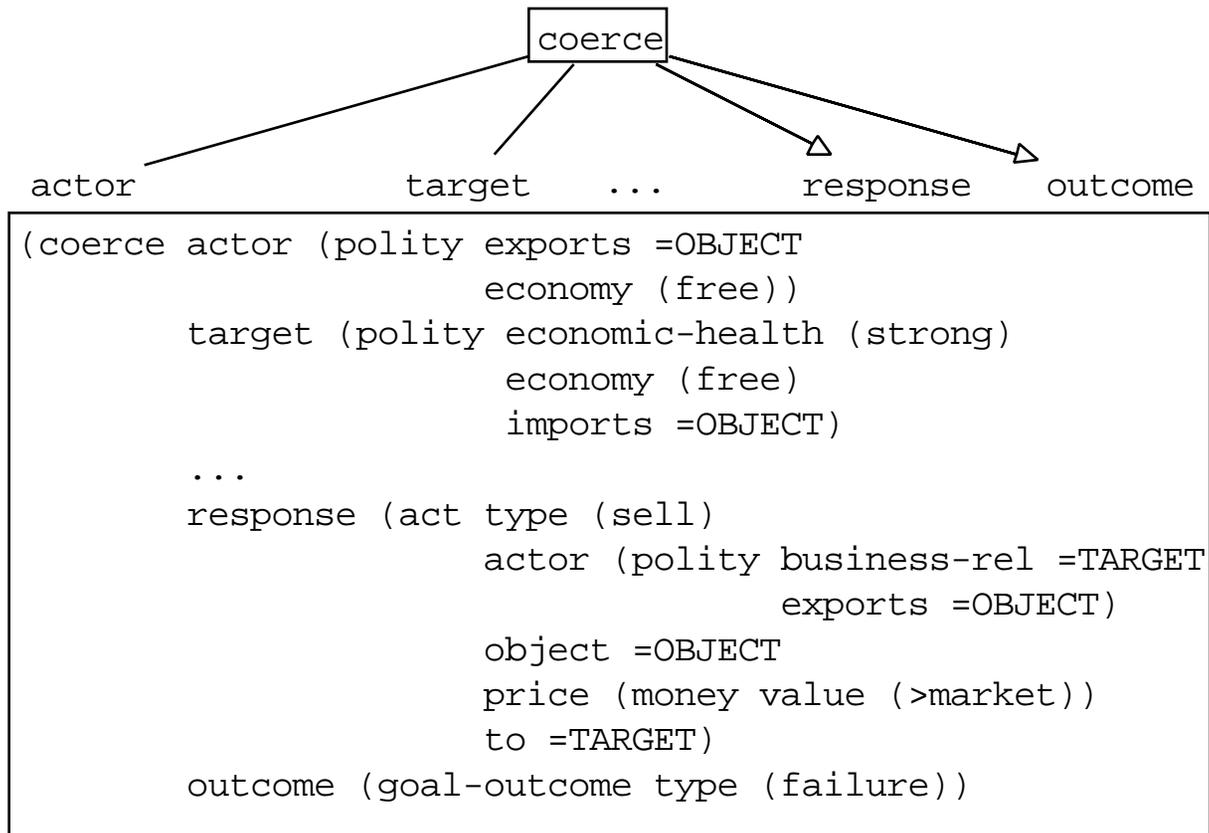
```
(state type (demand-increase)
  actor ?X: (polity economic-health (strong))
  object ?Y: (commodity availability (common)))
```

enables

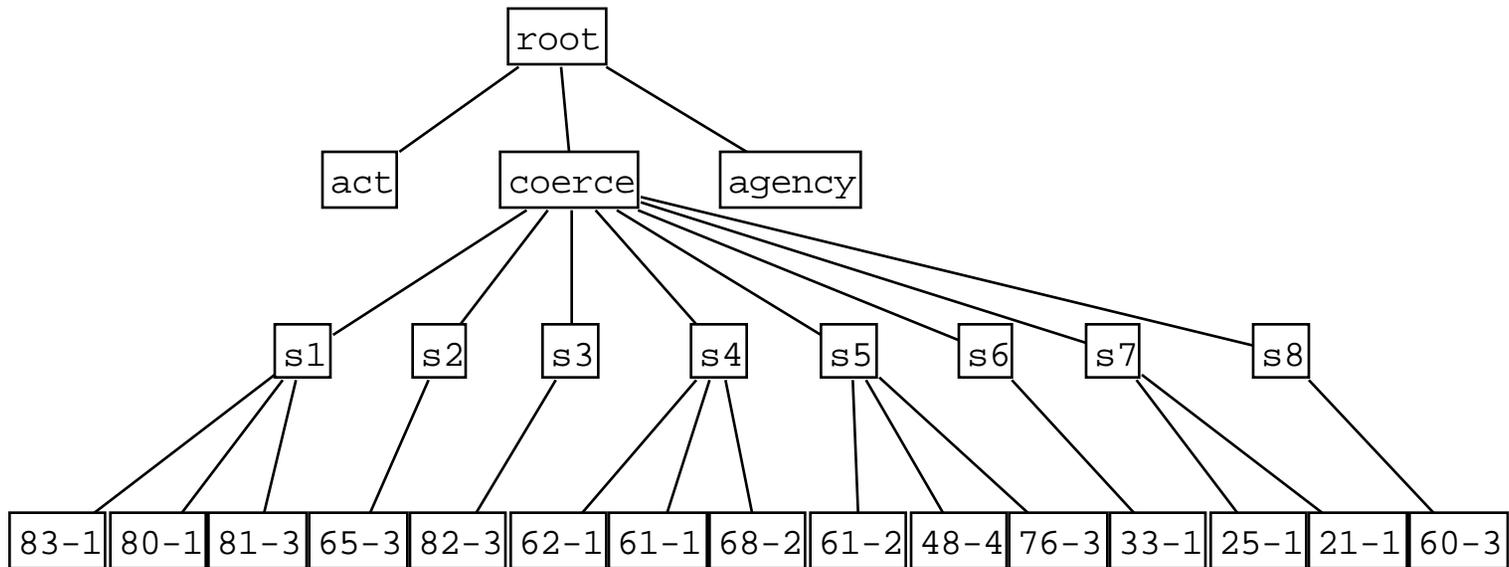
```
(act type (sell)
  actor (polity exports ?Y)
  to ?X
  object ?Y
  price (money value (>market)))
```

3. Purchase -> Possess

Indexing in memory by relevant features



Results after 15 cases



Conclusion

- Indexing explains desirable aspects of human memory retrieval
 - Retrieval of relevant exemplars by surface features
 - Encoding specificity (goal specific retrieval)
- Criticism by Waltz indexing caused by misunderstanding:
 - The “occasional, breakthrough insight” that requires more than 100 steps can result in a structure that is retrieved in less than a 100 steps.
 - Indexing by those surface features whose presence implies a deep feature constructs efficient recognizers.
 - Hierarchical memory search is not inherently serial
- Massively parallel memory searches (even without indexing) can make use of specially marked surface features at storage time.