Pathfinder

Pathfinder I (1992)

- Heckerman, Nathwani, and Horvitz
- Intended to raise general pathology diagnoses to the level of subspecialist pathologists.
- Integrated with a high-resolution laser disc of lymph pathology images
- Performance of Intellipath with a general pathologist exceeded that of the expert without Intellipath.

Intellipath

- Developed KBs for 24 of 39 areas of pathology
- ~2000-3000 systems sold.

Pathfinder

Lymph Pathology

- Lymph node removed from axilla, mediastinum

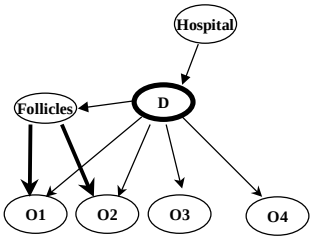
60 Diseases

- 25 Benign
- 9 Hodgkin's Lymphoma
- 18 Non-Hodgkin's Lymphoma
- 10 Metastatic Diseases

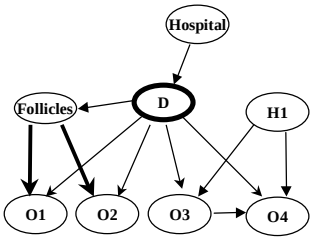
130 Features

Single Fault Assumption

Pathfinder I



Pathfinder II



Justification: Lymph diseases tend to invade only one area of a lymph node even when other diseases are present.

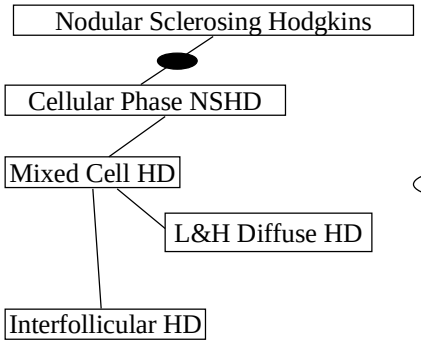
Similarity Networks

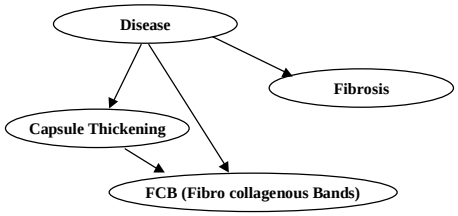
- Global belief network
  - Graph union of the local belief networks
- Context specific independence
  - f_i is independent of any disease d_k in D_{sub}
  - Omitting f from the local subset $D_{sub} = \{d_1, d_2, ...\}$ implies $P(d_k | f_i, D_{sub}) = P(d_k | D_{sub})$ for all d_k
  - This implies $P(f_i | d_k) = P(f_i | d_m)$ for all d_m, d_k, in D_{sub}
- Used to automatically construct partitions.
- Assessment time for both Pathfinder I and II
  - ~32 to 40 hours.

Similarity Networks

Problem:

m (binary) findings, n diseases: O(mn) assessments = 8000 assessments (naïve Bayes)

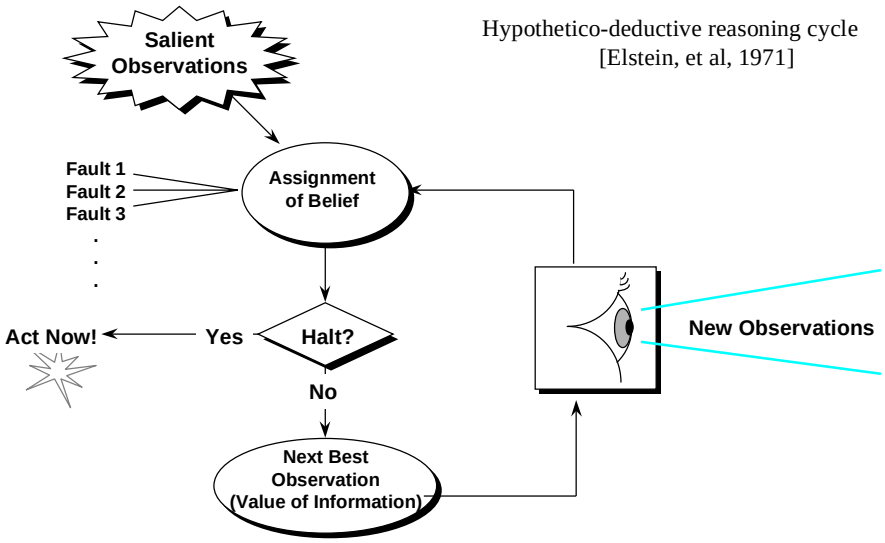




David Heckerman's dissertation, Similarity Networks, won the ACM best dissertation award

Diagnostic Reasoning

Hypothetico-deductive reasoning cycle
[Elstein, et al, 1971]

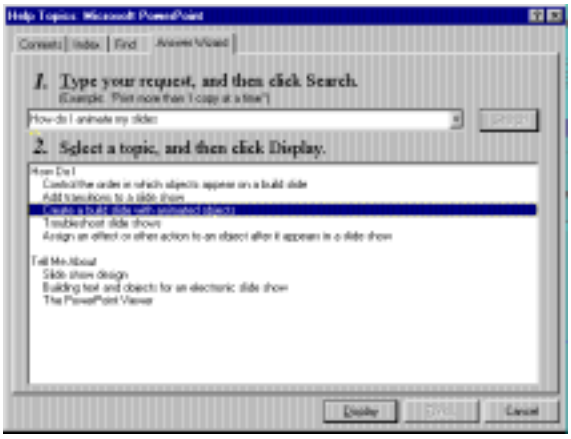


Next Best Finding

- Two models attempted
- Model I (used in paper and assessment)
  - Assess u(d1, d2) (paper)
  - Micromorts (10e-6 probability of death)
  - Problem “jumps around too much”
- Model II
  - Conditional KL distance driven.
    - Diseases are grouped into categories. (3/level)
  - Initially KL on disease groups (Pursue Groups)
  - When a disease hits a probability threshold, the system switches to conditional KL distance between the leading fault and remaining faults.

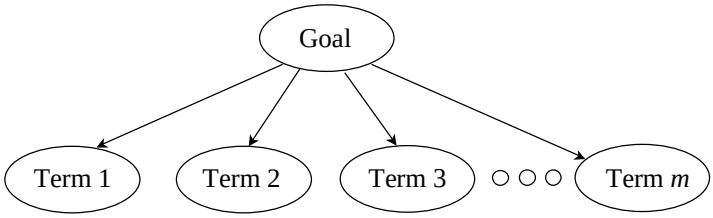
Answer Wizard (Microsoft)

David Heckerman and Eric Horvitz
Microsoft Research

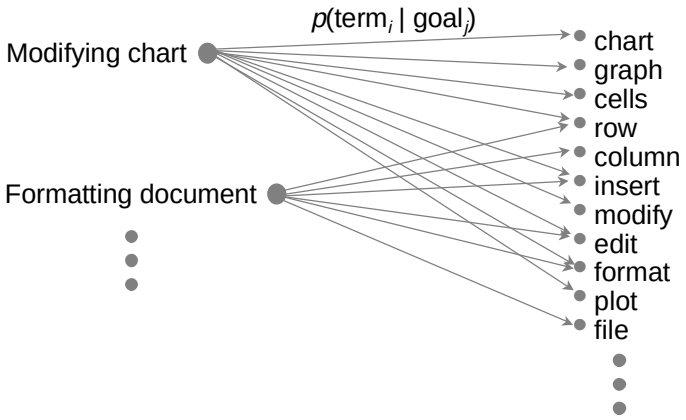


Bayesian term spotting

Only a single help topic is desired
The order of the query terms is unimportant
Only words/terms in a limited lexicon are used

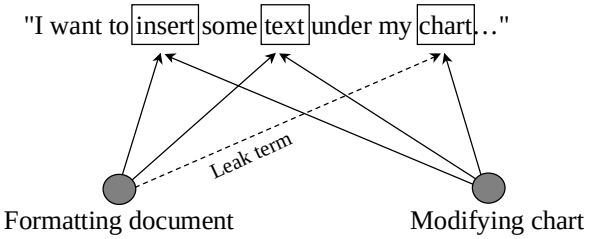


Probability assessment



probabilities assessed on a 1 to 13 log scale

Leak terms

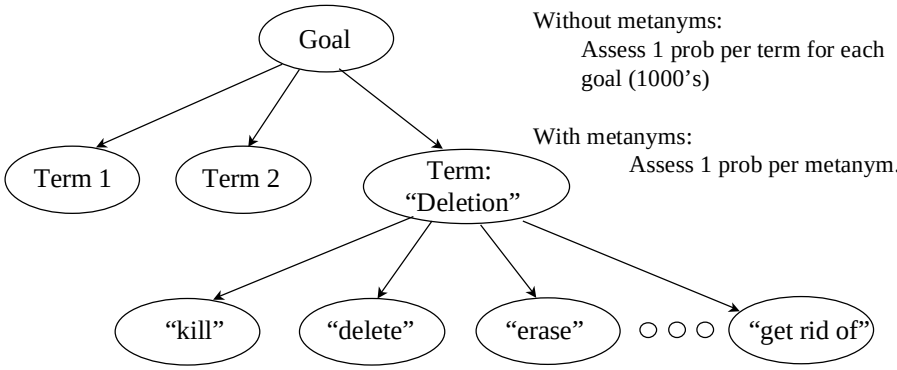


Metanyms

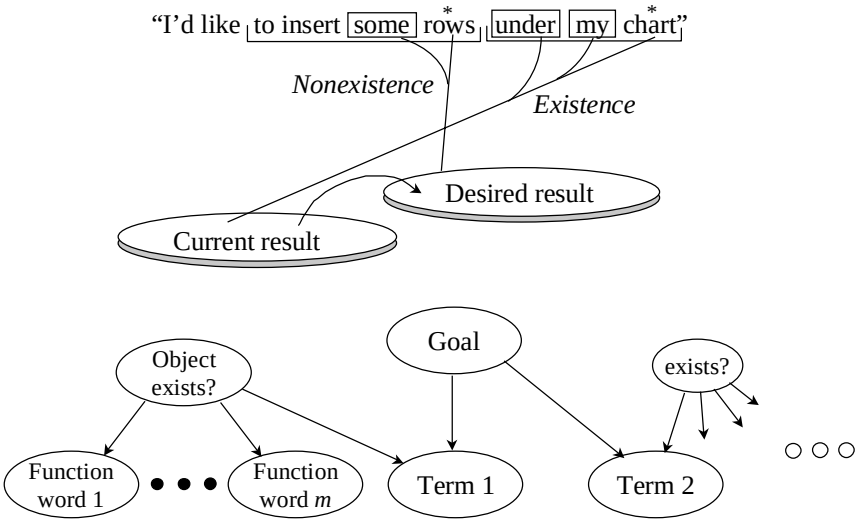
Reduces assessment time.
The major driver in the development of real systems.

Without metanyms:
Assess 1 prob per term for each goal (1000's)

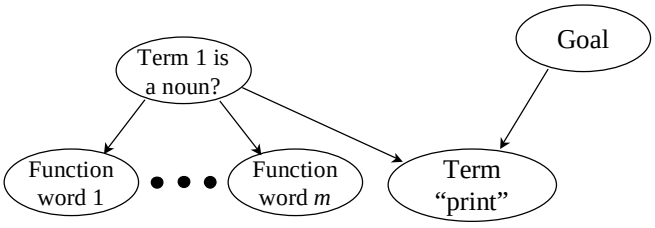
With metanyms:
Assess 1 prob per metanym.



Definiteness



Disambiguating Nouns and Verbs



How do I *print* this? (verb)
How do I make this *print* darker? (noun)

Evaluation

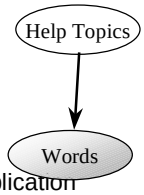
Sample queries obtained from online forums, email, MS usability lab

Users given tasks pictorially and asked to type in query to find the related help topic

75% of the time, correct topic was in top five (goal was 99%).

75% of those cases, user recognized the topic.

Microsoft Answer Wizard



Fielded in:

Office 95 Apps + MS Project

Size:

Approximately 1500 to 2000 cases (help topics) per application

Microsoft Word: > 1000 Topics, 5000 Terms, 145,000 Dependencies

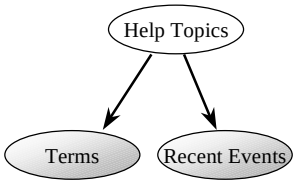
Model Construction:

Usability Experts: 40 Topics/week

Office Assistant (Office 97)

Extension of Answer Wizard

- tracks the user's past keyboard and mouse use (for example, menu searches and dialog box use.)
- events collected from 20K hours of usability experiments



Project

3/1 Project Proposal due.

Proposal: A 1 page technical abstract plus a bibliography and a schedule (milestones with dates).
You must schedule a meeting with me before your project will be approved.

3/26 Progress Report I.

Progress Report: 1-2 page report on status including early results and milestones achieved. Alternatively, you may schedule a meeting with me to discuss your progress.

Project

4/9 Progress Report II.

Progress Report: 1-2 page report on status including results and milestones achieved. You should have completed most of your project by this point. Alternatively, you may schedule a meeting with me to discuss your progress.

4/21 - 4/23 Project Presentations

Each student will have 25 minutes to present their project. This will be the presentation order (randomly selected):

4/21: Maria, Noam
4/23: Mark, Liang

4/23 Final report due.