Distributing Knowledge
in Composite Information Retrieval Tasks

Leslie L. Daigle

Vice President Research

Bunyip Information Systems Inc

Montréal, Canada

Patrik Fältström

Senior Researcher

Tele2

Stockholm, Sweden

October 31, 1996

What is “search”?

- We have
  - information need
  - information source
- We want to
  - match description of need against source
- With some weighting of
  - maximized coverage
  - minimized inclusion of spurious data

What is “Internet search”?

- As above + distributed ownership
- Heteregeneous
  - No global control
  - Limitations on access
  - Multiplicity of languages, formats, implications...

⇛Often need *composition* of multiple result sets

Different kinds of data

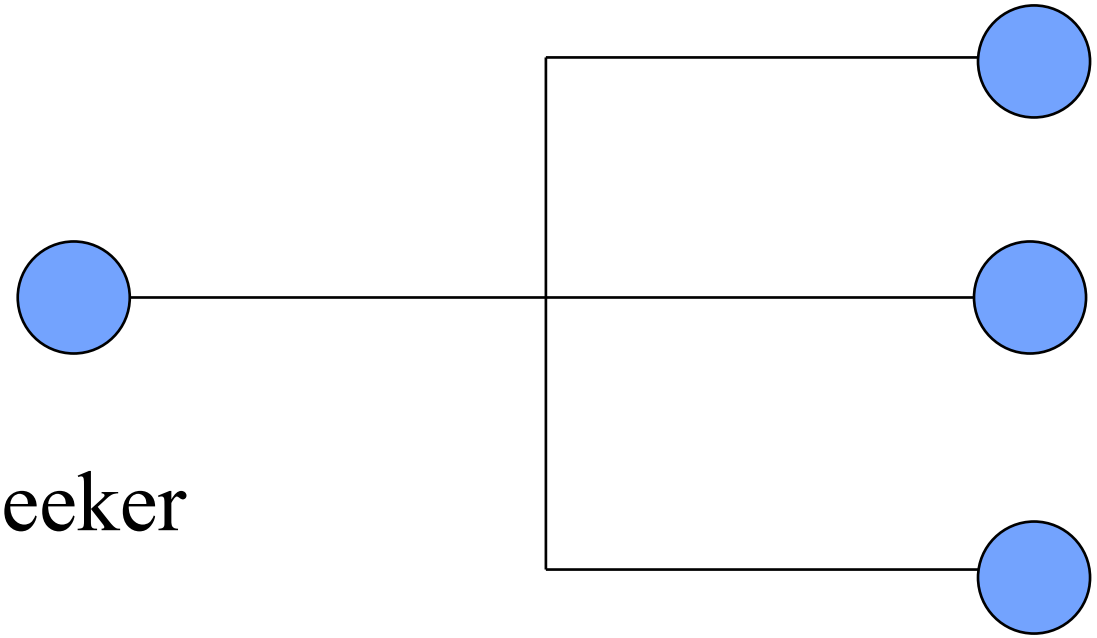
- Different **owners** of
  - Source
  - Metadata
  - Index
  - Query
  - Client's knowledge

Who are the players?

- Producer
  - Owner of information
- Indexer
  - Mediator or “middle man”
- Consumer
  - Seeker of information

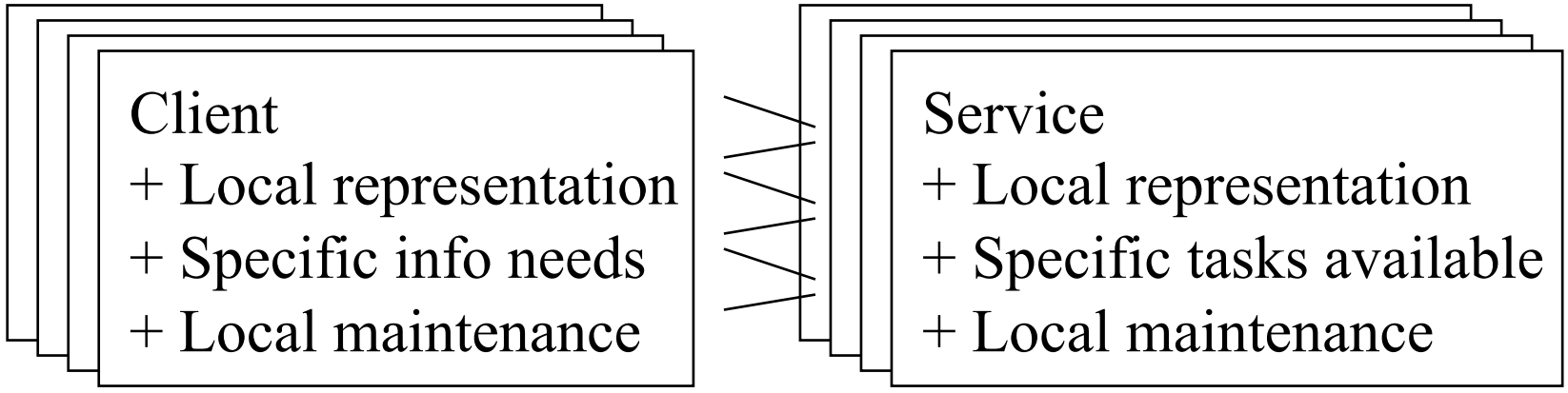
Knowledge distribution model

- Who owns what of the above?



Abstract query transaction mechanics

- Description of the mechanics
  - Chaining
  - Referrals
  - What is a client and a service?



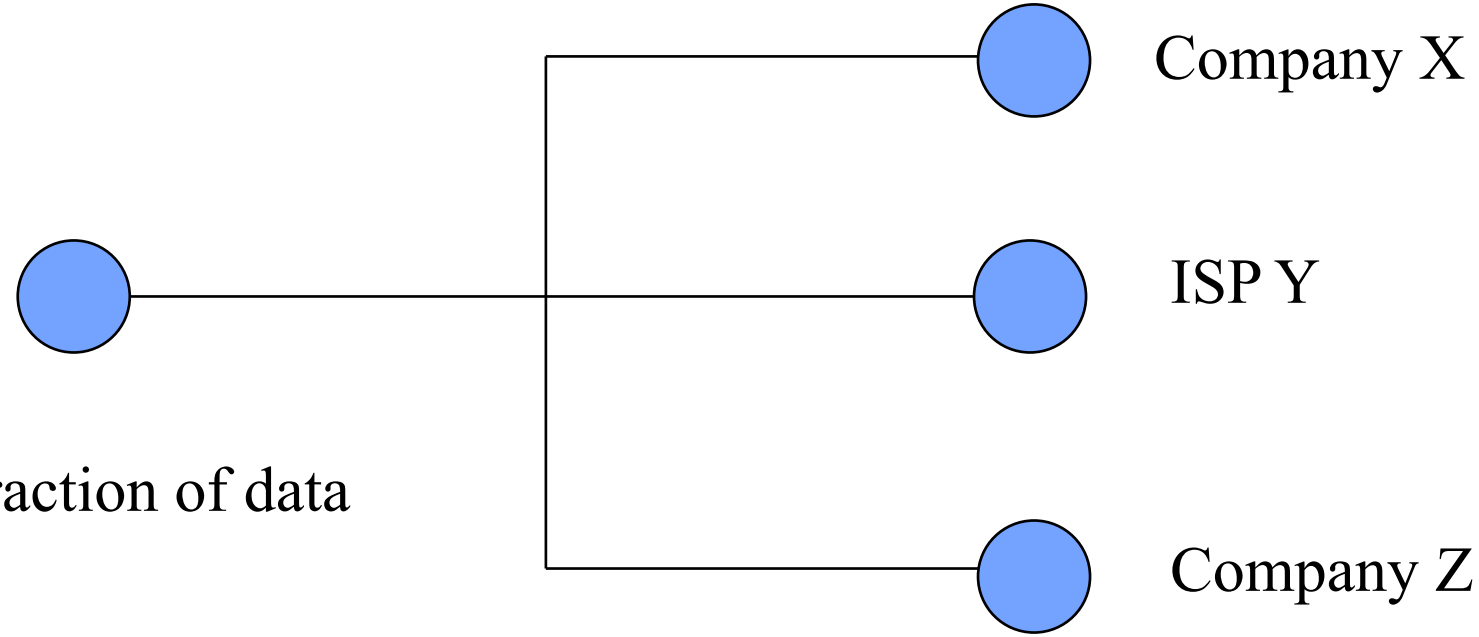
Example 1 (White-pages)

- “Finding people on the net”
- A global problem, requiring a global solution

White pages data

- Source
  - (Company) personnel files
- Metadata
  - Descriptors
- Index
  - Whois++, LDAP
- Query
  - Name, Email address, fragments of (misspelled?)
- Client's knowledge
  - Person sought, type of person sought

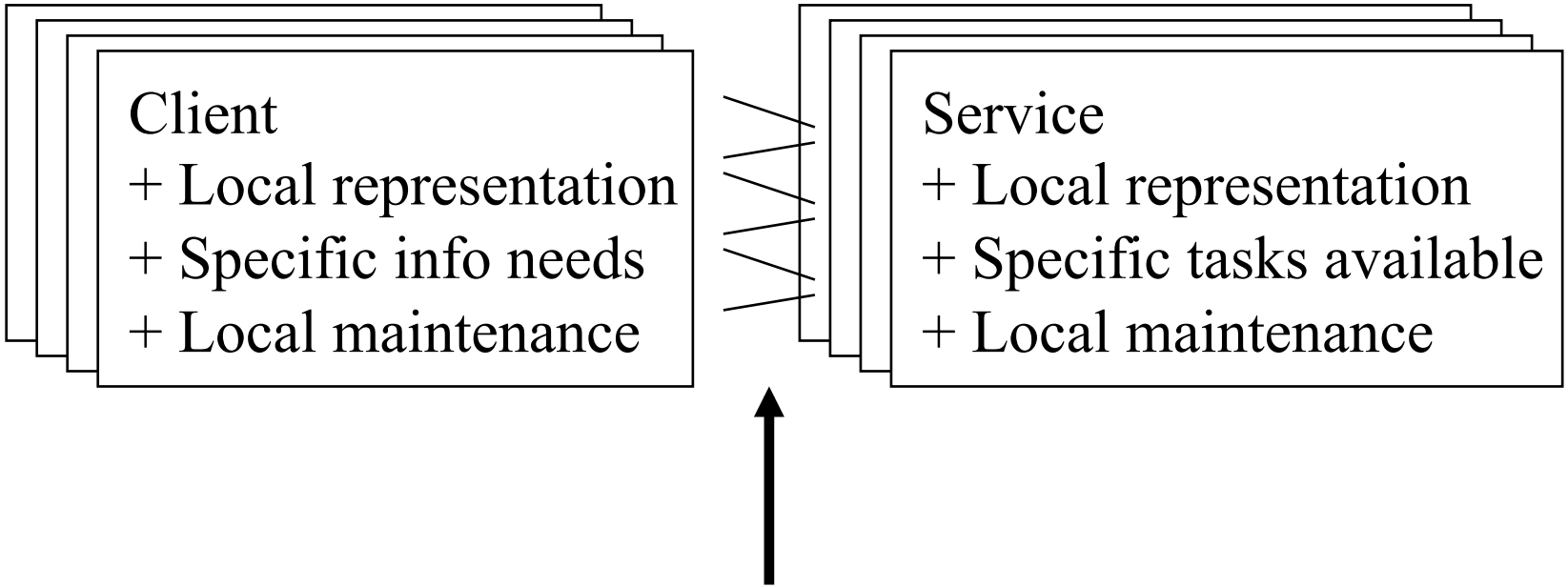
Knowledge distribution model with White pages data



Query: extraction of data from user

Different sources, technologies, cultural conventions.
Can match “similar names”, use language constraints etc.

Abstract query transaction mechanics



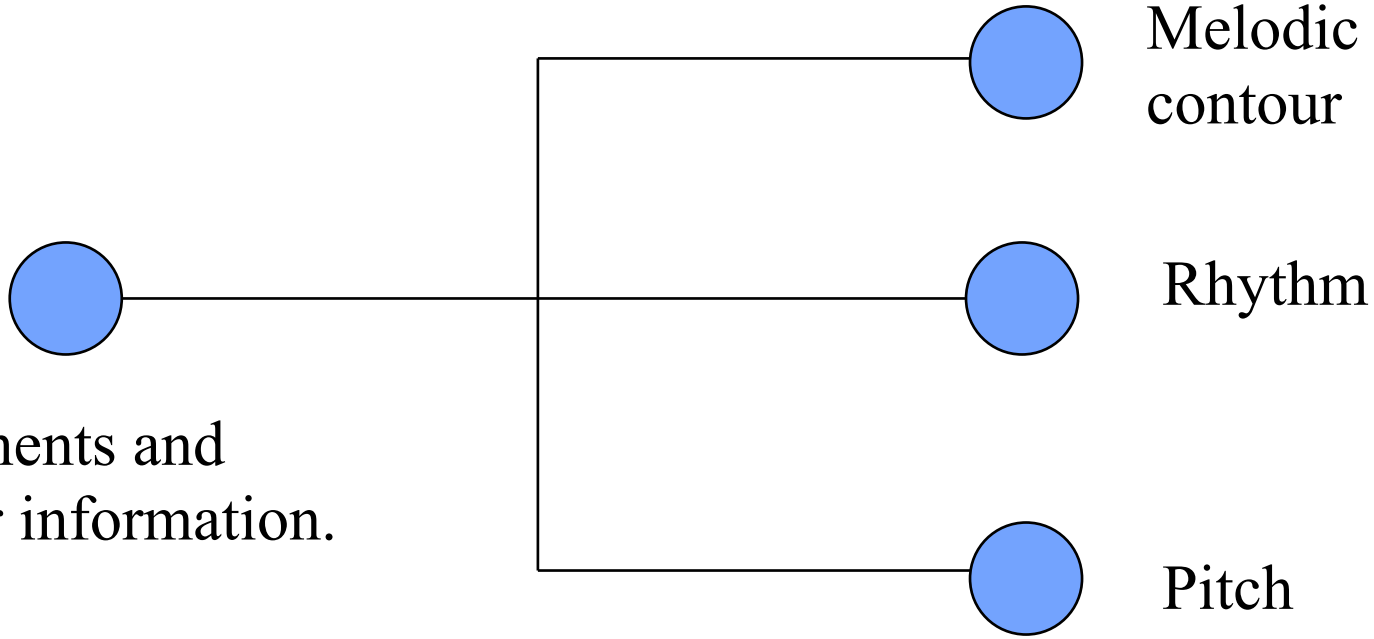
Example 2 (Music)

- Finding music by example
  - Not particular rendition
  - The piece in general
- Can be handled by splitting into 3 dimensions
  - Melodic Contour
  - Rhythm
  - Pitch

Music data

- Source
  - Music (performance, recording or transcription)
- Metadata
  - Melodic contour, rhythm pattern etc
- Index
  - Matcher of representations - different for each dimension
- Query
  - Fragment
- Client's knowledge
  - Perhaps includes more than performance fragment

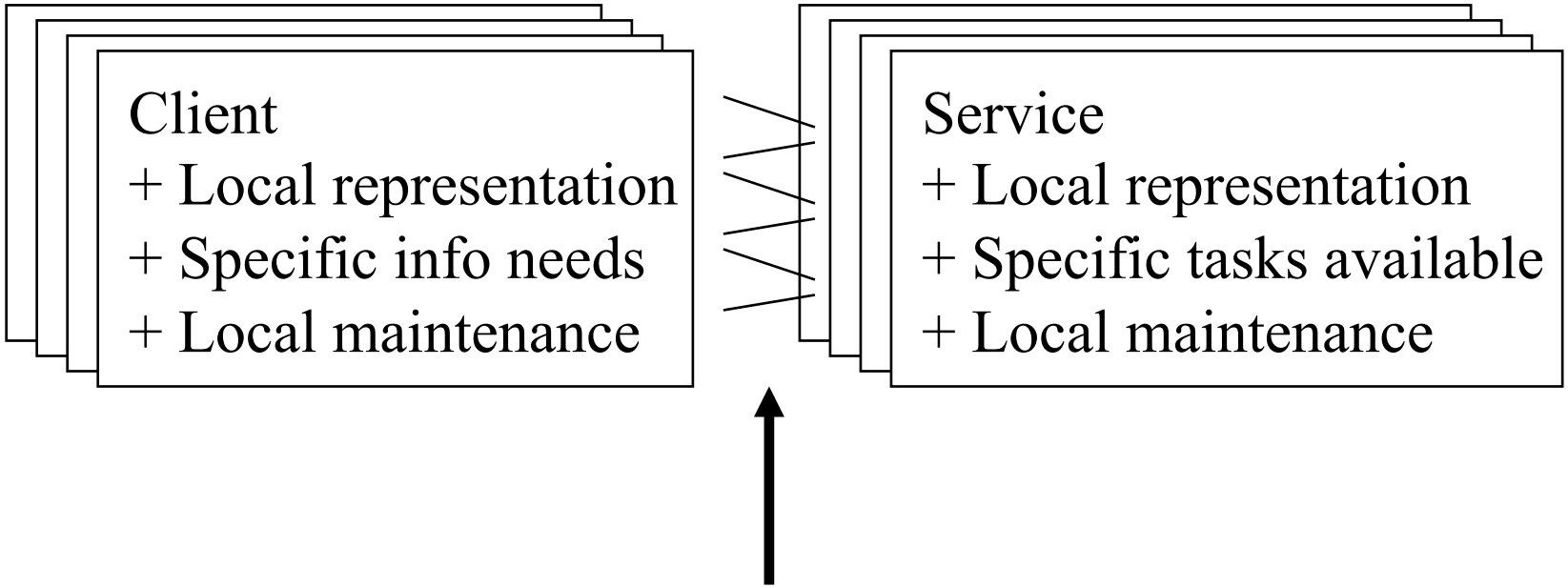
Knowledge distribution model with music



Query fragments and maybe other information.

Can pick which dimension to weight most appropriately.

Abstract query transaction mechanics

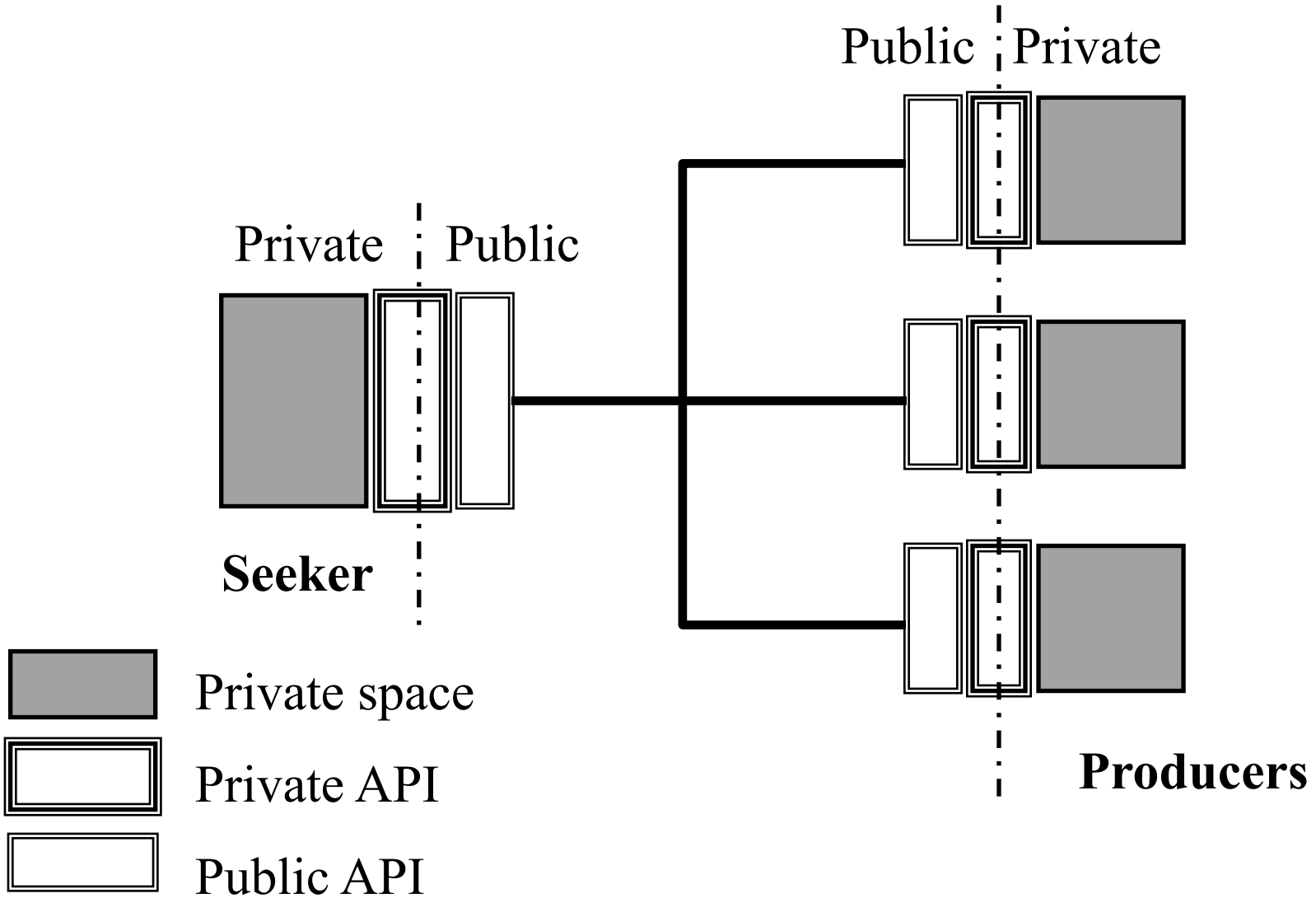


Need to know how to analyse
search fragment and
formulate query + responses.

Example that doesn't work...

- Roaming agents and information ownership
  - Source – Owned by document owner
  - Metadata – Owned by the index owner
  - Index – Owned by the index owner
  - Query – Owned by the user
  - Client's knowledge - Owned by the user
- An agent is a user representative
  - Should only play with user-owned information

Privacy



Roaming agents

- An agent can not have an expertise in the source or any intermediate results.
- The expertise must stay as close as possible to the owner of the information at each stage in the process of a search.
- Bandwidth is not a problem (basically).
- More sophisticated global policies & expression of searches needed!

Conclusion

- Knowledge must be present at all stages in the process of a search – from the Seeker to the Producer and back
- It is impossible to replace missing knowledge at one step in the process by knowledge at some other stage – only guesses are possible
- Examples.....next Thursday...

Next talk

- A practical demonstration on how knowledge is missing or missused in today's systems (Lycos, AltaVista, X.500, URL, PH)
- A practical demonstration on how knowledge can be distributed (Whois++, Archie, CIP, URA, URN, DNS)