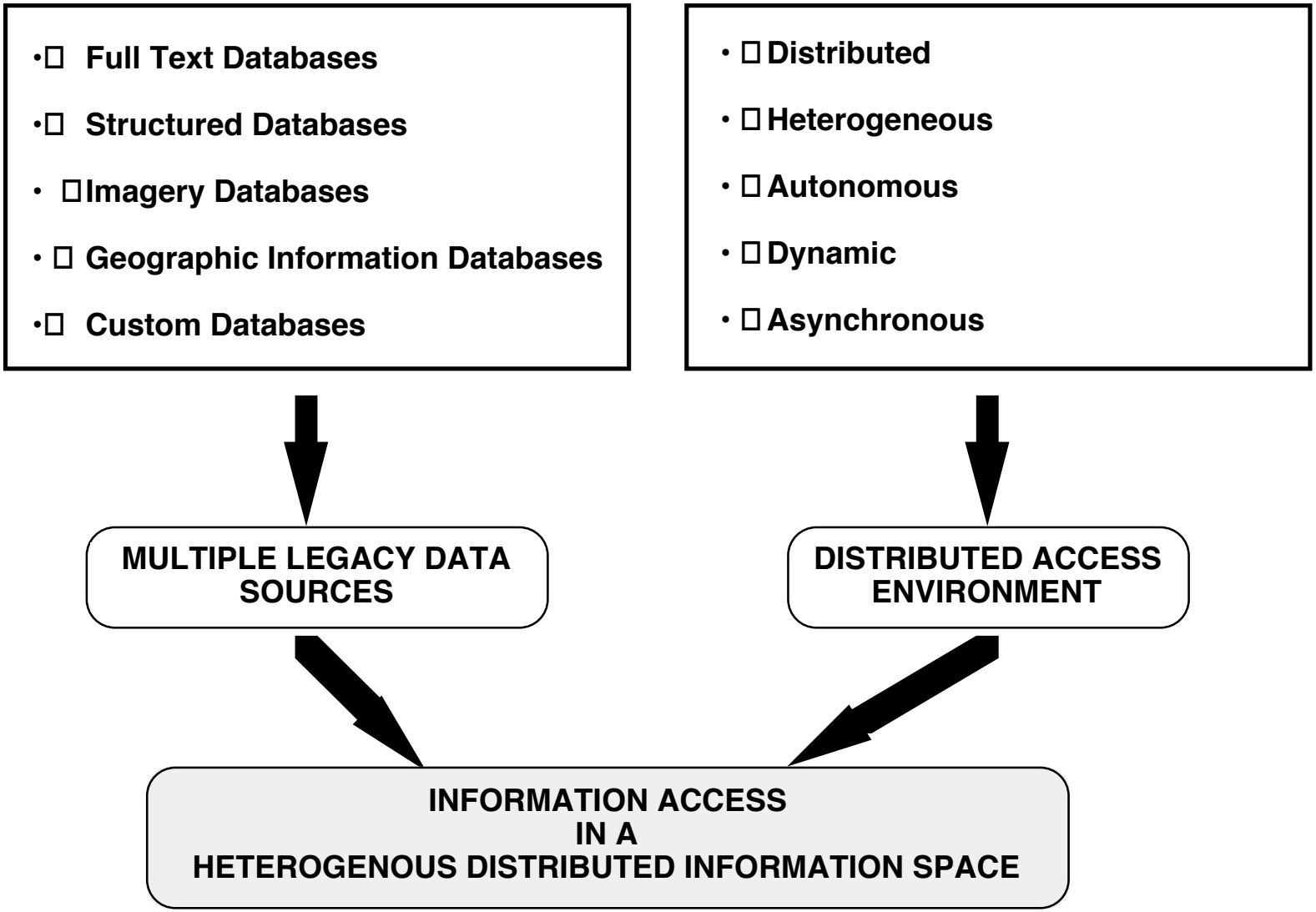
MINERVA
A System for Heterogeneous Information Access

Richard M. Tong, Ph.D.
Senior Associate
Booz•Allen & Hamilton, Inc.
Mountain View, CA

Presentation Outline

- Problem Statement and Project Goals
- MINERVA Infrastructure
- Concept-Based Retrieval for Full-Text
- Extensions to Relational Data
- Source Selection
- Future Research
- Summary and Conclusions

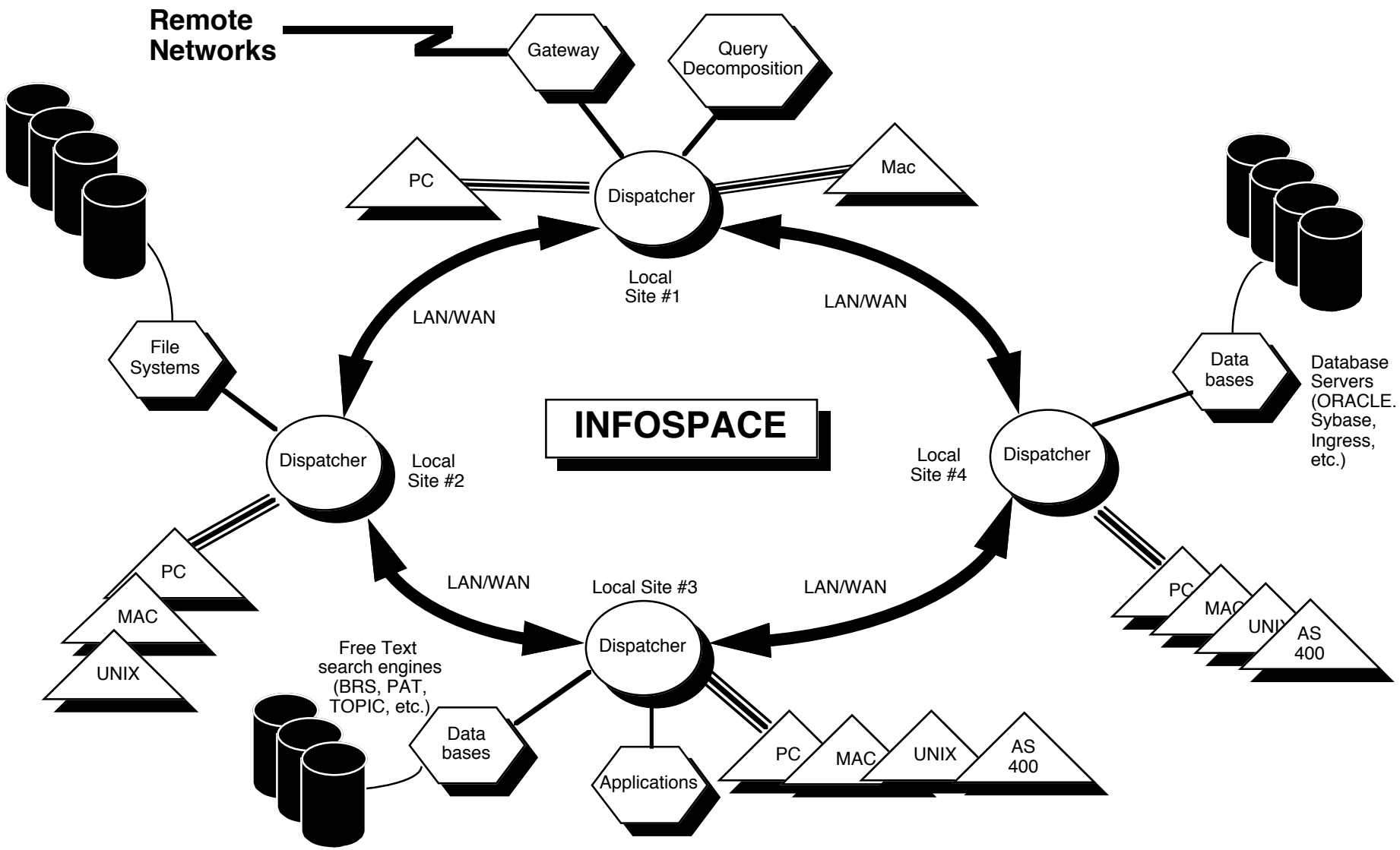
The Information Access Problem



MINERVA Goals

- Provide a distributed information operating environment that can work with multiple hardware platforms and multiple legacy data sources.
- Provide an open-architecture that can support multiple information access “engines” and information analysis tools.
- Provide the end-user with a single, transparent point of access to the heterogeneous distributed information space.

Distributed Architecture



Distributed Services Control Strategy

- User actions at Desktop requests services from Dispatcher
- If local services can provide service, Dispatcher provides to Desktop; Dispatcher also sends service request around ring
- Each Dispatcher around ring notifies requesting Dispatcher of its ability to provide service and then provides results
- When requesting Dispatcher receives results from services, they are presented to user via Desktop
- Timeouts, ring reconfiguration, knowledge of services, etc. are provided in **Distributed Information Operating Environment**

The Meta-Language

- **The MINERVA meta-language represents both requests for information and the associated responses**
- **A meta-langauge request/response is a data packet transported via the underlying network protocol (e.g., TCP/IP)**
- **The meta-language is extensible and designed to be a super-set of existing protocols (e.g., Z39.50)**
- **The current implementation builds "sequences" from primitives that include:**

  Request line: describes the purpose of the sequence

  Destination line: lists all the destinations that the packet should be sent to

  RespondTo line: lists all the destinations that need to receive responses

  MultiMedia line: information about multi-media filenames

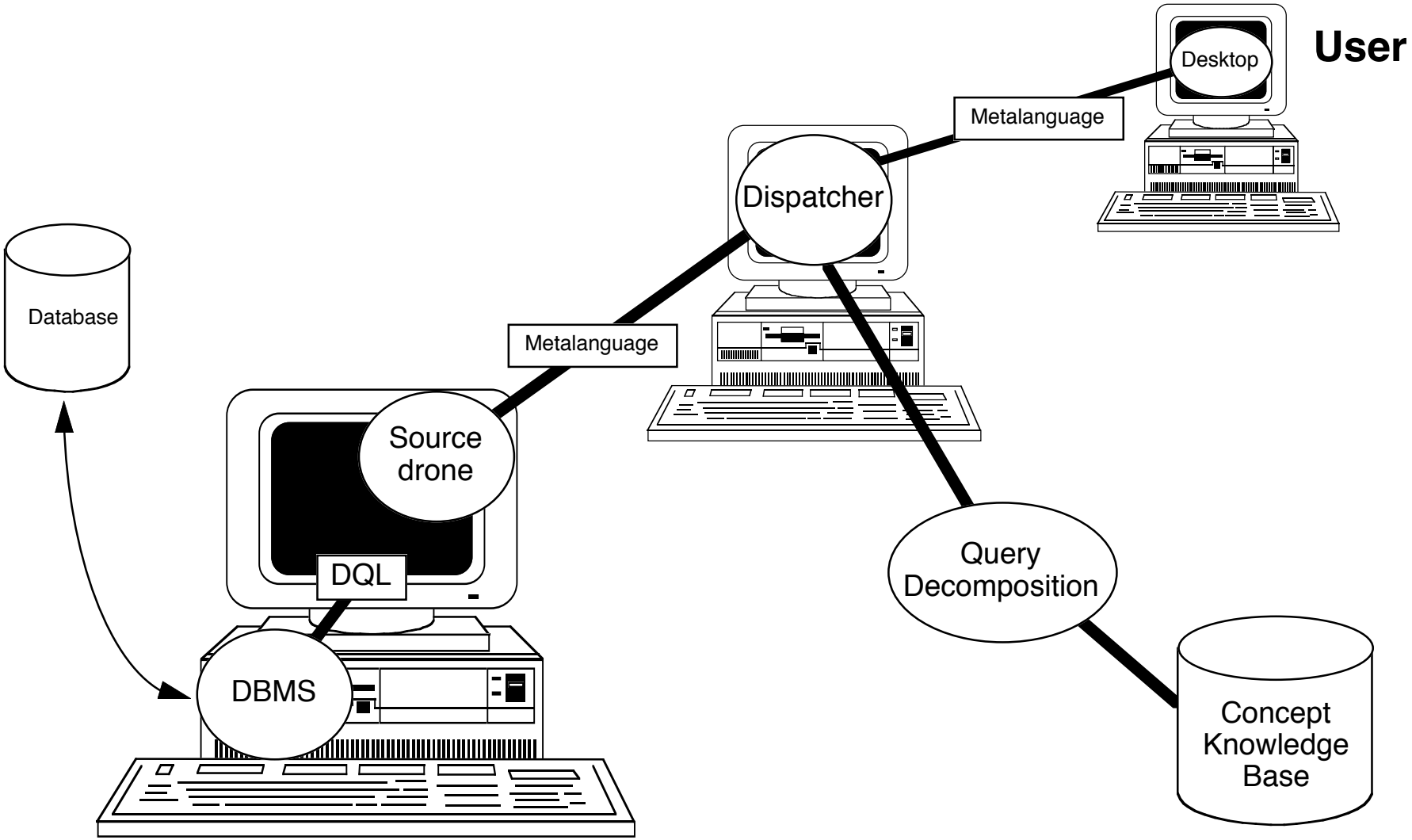
  Query line: query formulated in MINERVA's query reference language

  FormattedData line: used to convey administrative information about responses

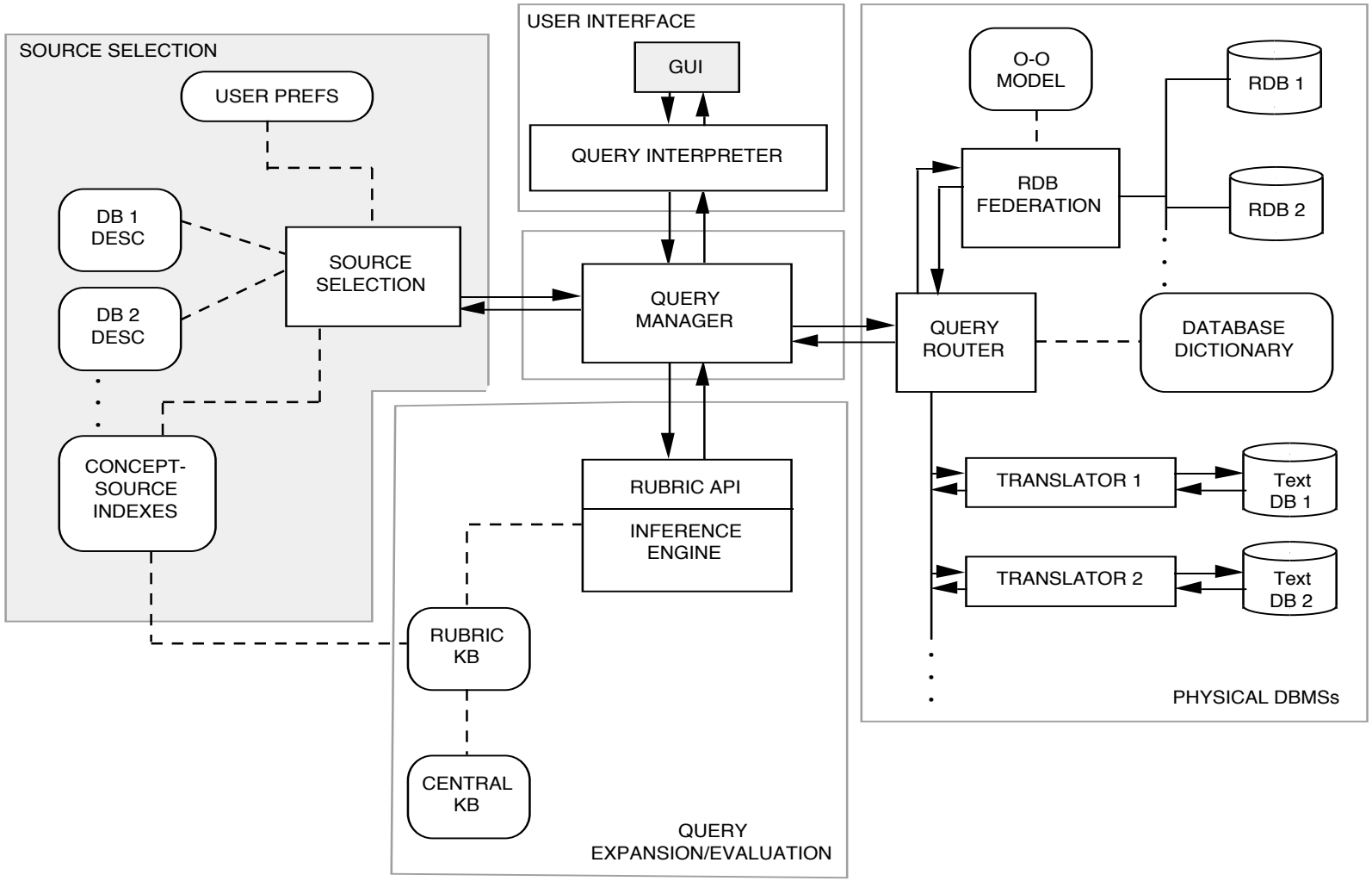
  Parameter line: used to convey miscellaneous/user defined control information

  Done line: marks the end of the sequence

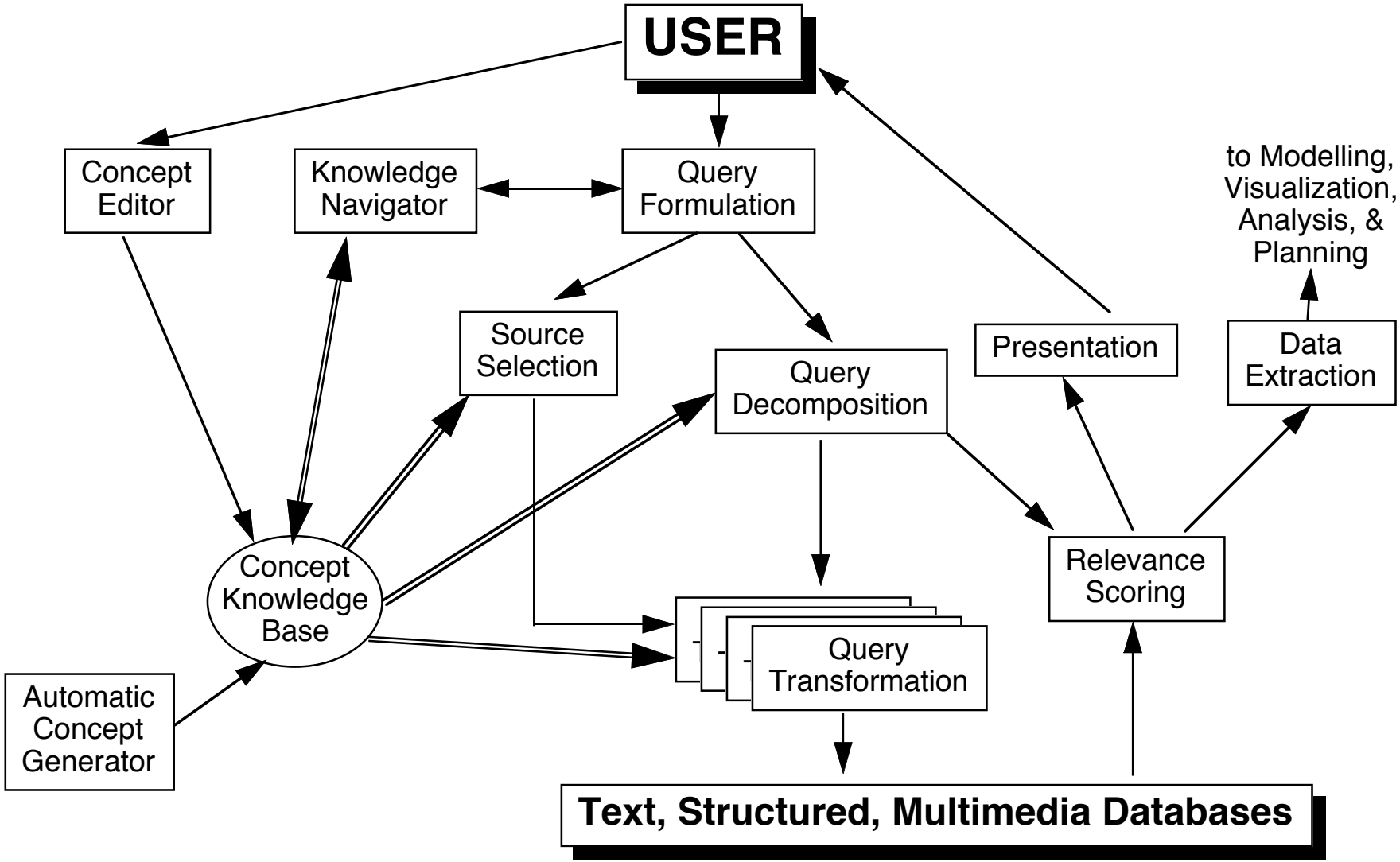
Access Example



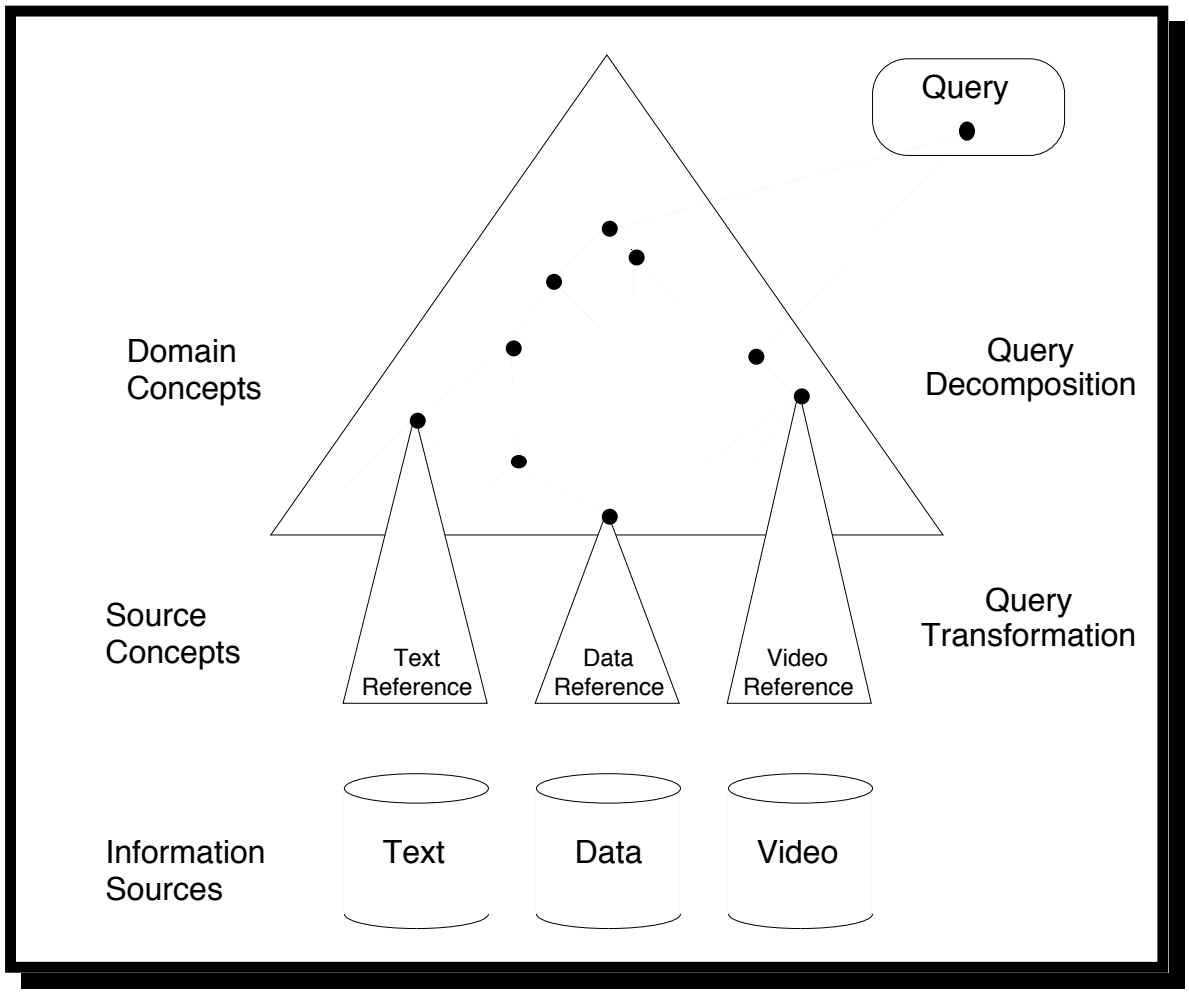
MINERVA Functional Elements



Search and Retrieval Flow in MINERVA



Concept-Based Access Approach



- Queries are combinations of domain concepts
- High-level domain concepts are successively mapped into lower level domain concepts
- Low-level domain concepts are mapped into references in each specific medium in the information sources
- Decomposition weights provide relevance scoring mechanism

What is a Concept?

- A concept is a domain dependent entity that carries the semantics of the domain and that can be described and organized without reference to the specific document collections or data sources

- Naming a concept corresponds to distinguishing a sub-class of a larger class of entities

- A collection of semantically related concepts defines a “retrieval model” for the domain

- Concept definitions are relative rather than absolute

Concept Definitions

```
(ATTRIBUTE meeting_event :action)
(ATTRIBUTE meeting_event :actors)
(ATTRIBUTE meeting_event :topic)
(ATTRIBUTE meeting_event :location)

(DEFINES :location (*OR* moscow washington vienna)

(SUBSET meeting_event diplomatic_meeting)
(SUBSET diplomatic_meeting us_soviet_summit)
(SUBSET meeting_event white_house_meeting)

(DEFINES white_house_meeting:location white_house)
```

Semantic Structure

What is Evidence?

- Evidence is any feature that can be used to determine whether a document is about a domain concept
- Evidence is usually derived from the content of the document but can also be derived from the logical structure of the document
- Evidence is the link between the document collection and the retrieval model

Evidential Structure

```
(*OR* "KREMLIN" "MOSCOW" "RUSSIA")
(*AND* "REAGAN" (*OR* "MOSCOW" "BEIJING"))
(*NOT* (*AND* "REAGAN" "MOSCOW")
(NEAR "PRESIDENT" "REAGAN")
(SENTENCE ""GORBACHEV" "REAGAN")
(PRECEDES "SINO" "SOVIET")

(SOURCE wall_street_journal)
(TYPE feature_article)

(EVIDENCE moscow (*OR* "KREMLIN" "MOSCOW" "RUSSIA"))

(SCOPE us_euro_meeting paragraph)
```

Representation of Belief

Every concept