Agentized, Contextualized Filters for Information Management

AAAI Spring Symposium

Agent-Mediated Knowledge Management

Stanford University

David A. Evans, Gregory Grefenstette, Yan Qu, James G. Shanahan, Victor M. Sheftel

Clairvoyance Corporation

March 25, 2003

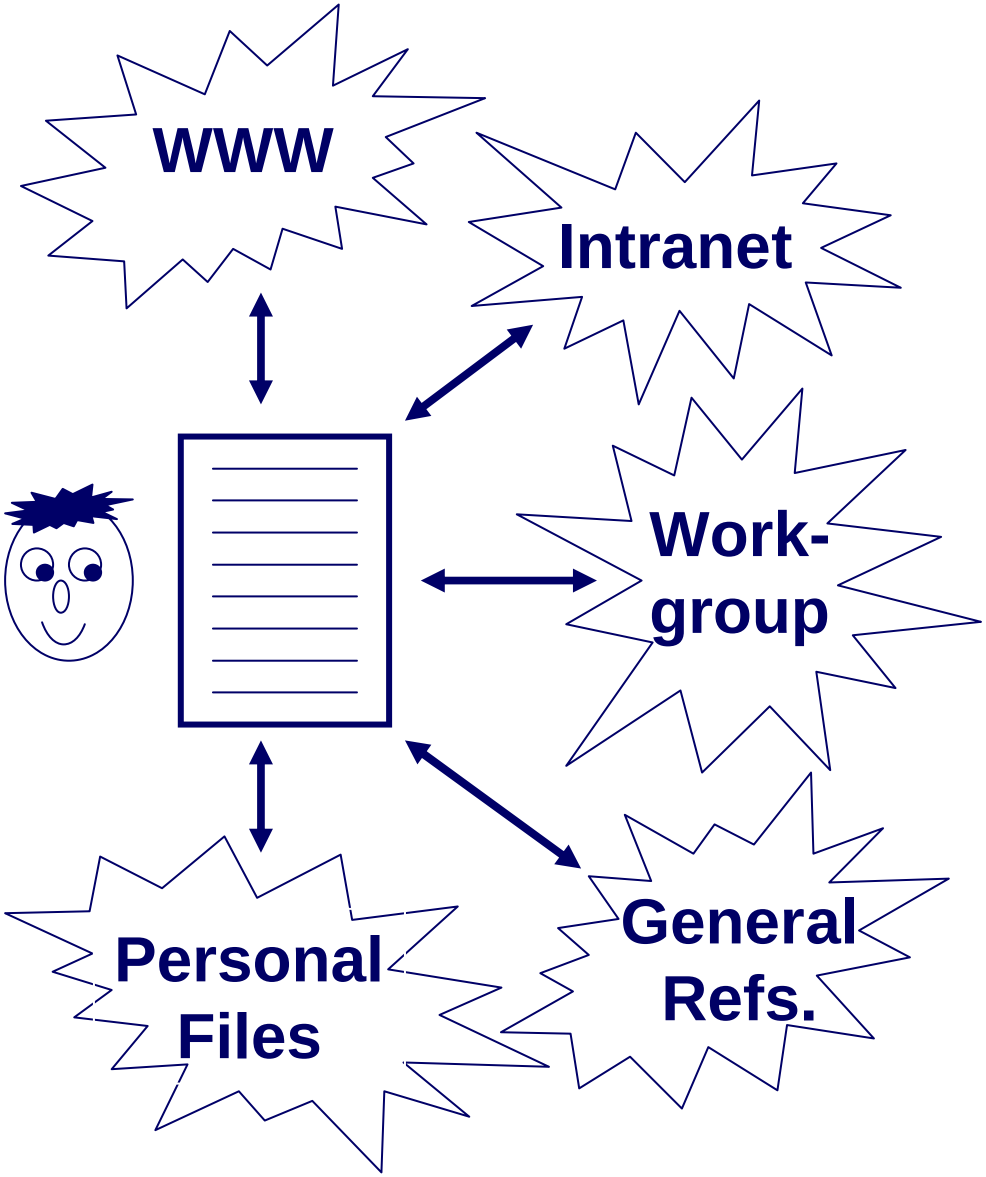
Overview

- Document-Centric Systems
- ViviDocs, IOs, ACFs
- Model
- Operational Screenshots
- Conclusions

Overview

- **Document-Centric Systems**
- **ViviDocs, IOs, ACFs**
- **Model**
- **Operational Screenshots**
- **Conclusions**

Document as Information Hub



- **Document is linked to *relevant* information**
- **Document is a *lens* focusing information on the user’s interests, anticipating user’s needs**

Document-Centric Systems

- **Typed Entity Recognition**
  - Niche browsers
    - Flipdog [Monster]
    - Citeseer [NEC]
- **Exploiting User Context**
  - Modifying search using accumulated context
    - Remembrance Agent [Rhodes & Maes 2000, MIT]
  - Contextual Search
    - Watson [Budzick & Hammond 2000]
- **Periodic, Anticipatory Retrieval**
  - Background
    - Document Souls [Shanahan & Grefenstette 1999, Xerox]
    - Kenjin 2000 [Autonomy]

Overview

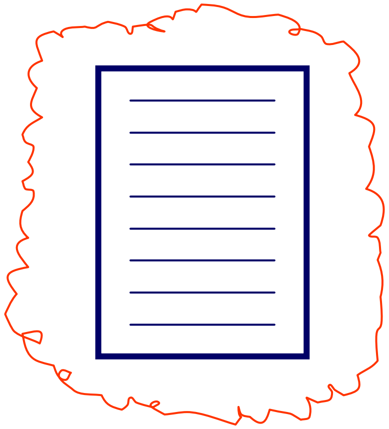
- Document-Centric Systems
- ViviDocs, IOs, ACFs
- Model
- Operational Screenshots
- Conclusions

ViviDocs

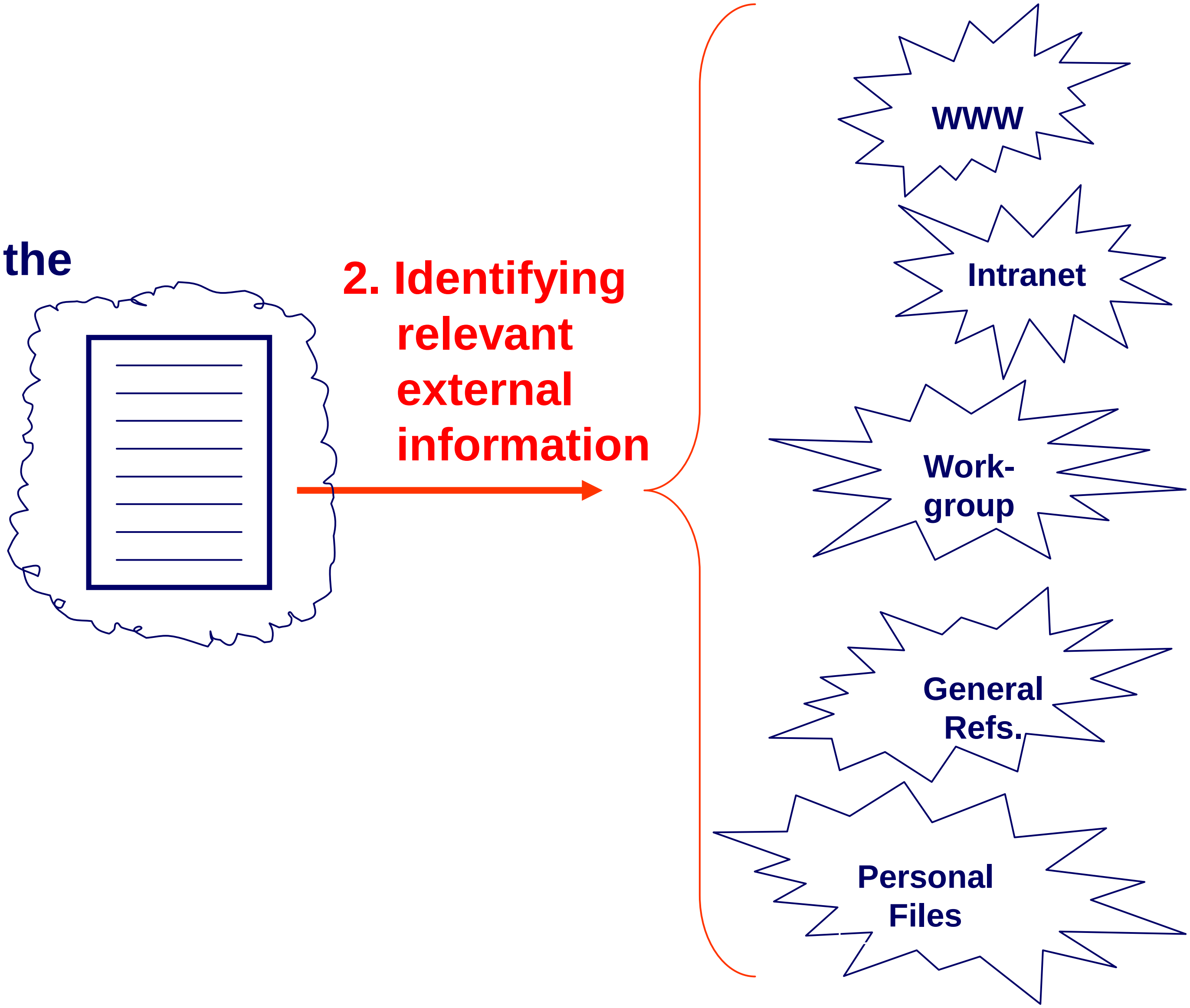
- **“Living Documents”**
- **Formalizing, generalizing possible relations between document and external data**
- **Fusion of IR, ML, NLP, user modeling**
  - adaptive filtering
  - question answering
  - classification
  - entity recognition
  - relevance feedback

ViviDocs Challenges

1. **“Parsing” the document**



ViviDocs Challenges



1. “Parsing” the document

2. Identifying relevant external information

WWW

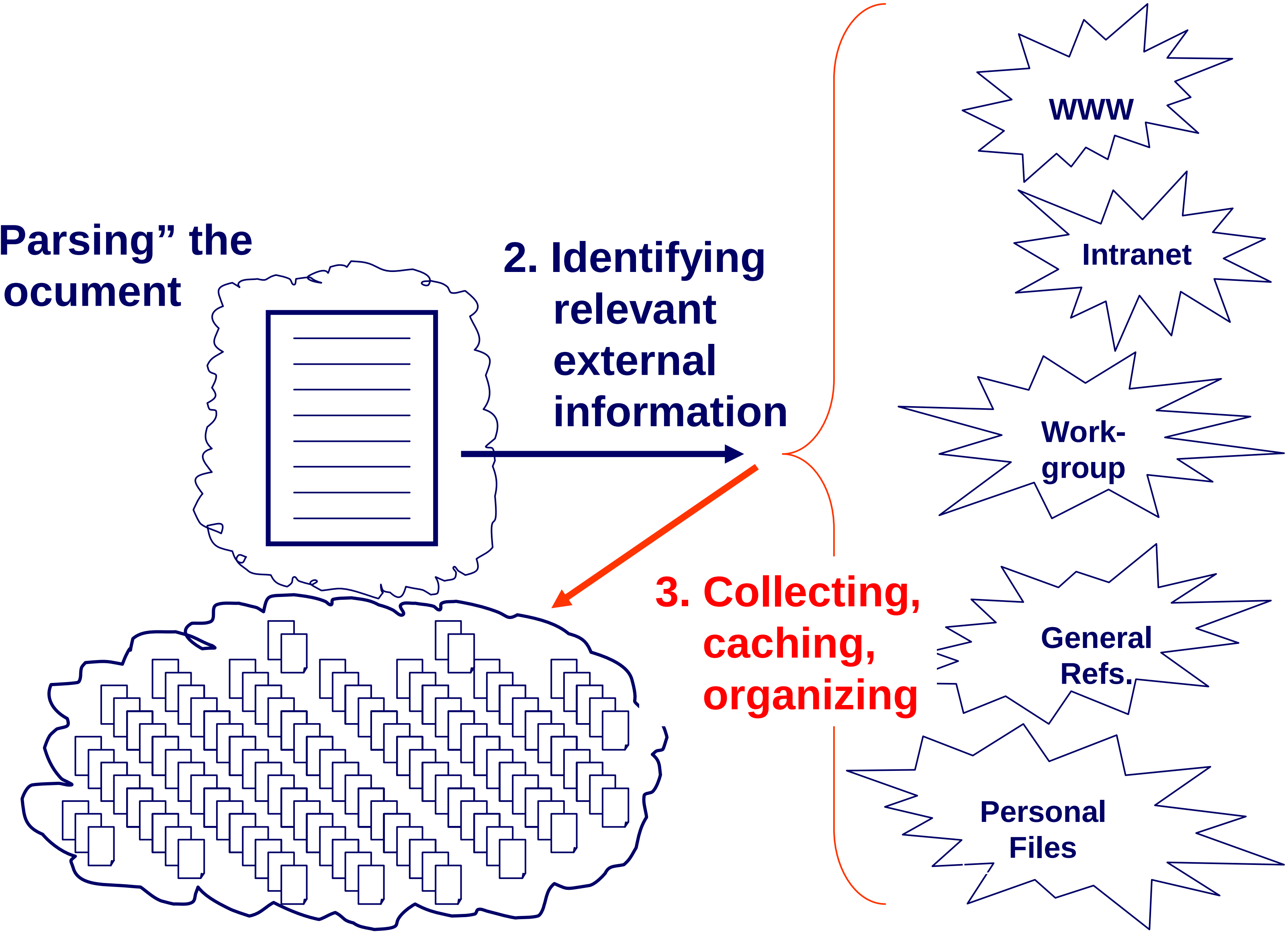
Intranet

Work-group

General Refs.

Personal Files

ViviDocs Challenges



1. “Parsing” the document

2. Identifying relevant external information

3. Collecting, caching, organizing

WWW

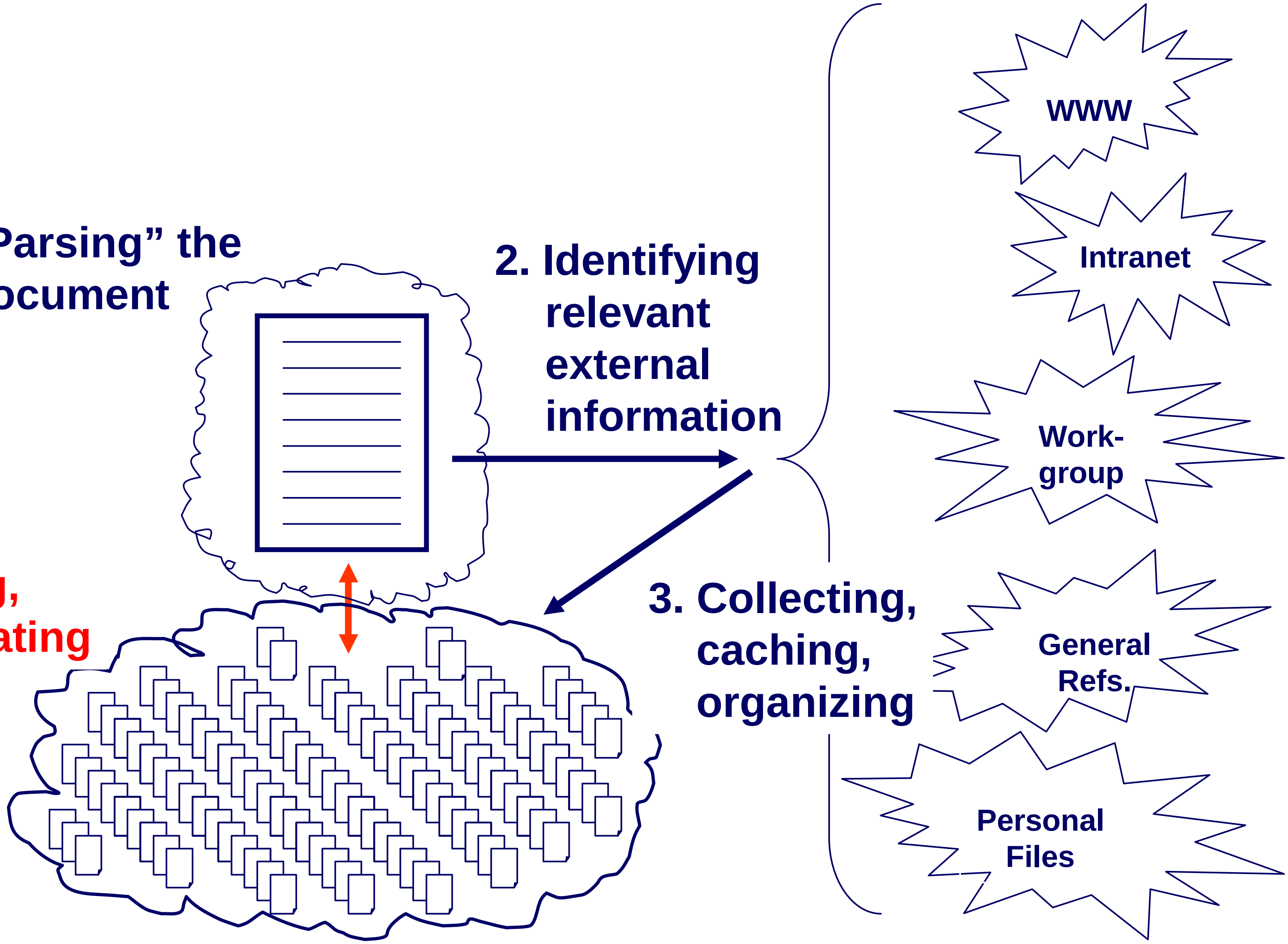
Intranet

Work-group

General Refs.

Personal Files

ViviDocs Challenges



1. “Parsing” the document

2. Identifying relevant external information

3. Collecting, caching, organizing

4. Linking, associating

WWW

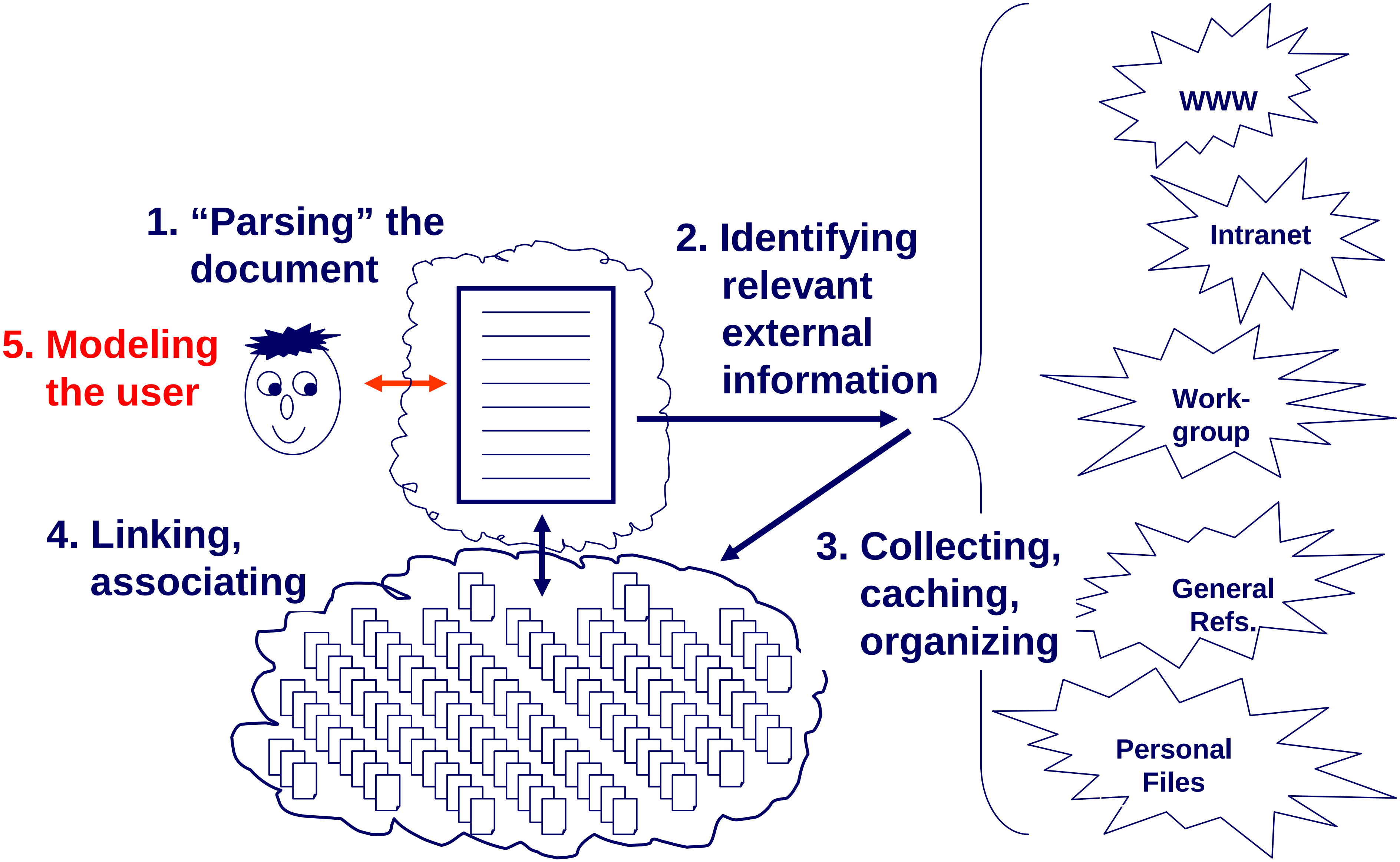
Intranet

Work-group

General Refs.

Personal Files

ViviDocs Challenges



1. “Parsing” the document

2. Identifying relevant external information

3. Collecting, caching, organizing

4. Linking, associating

5. Modeling the user

WWW

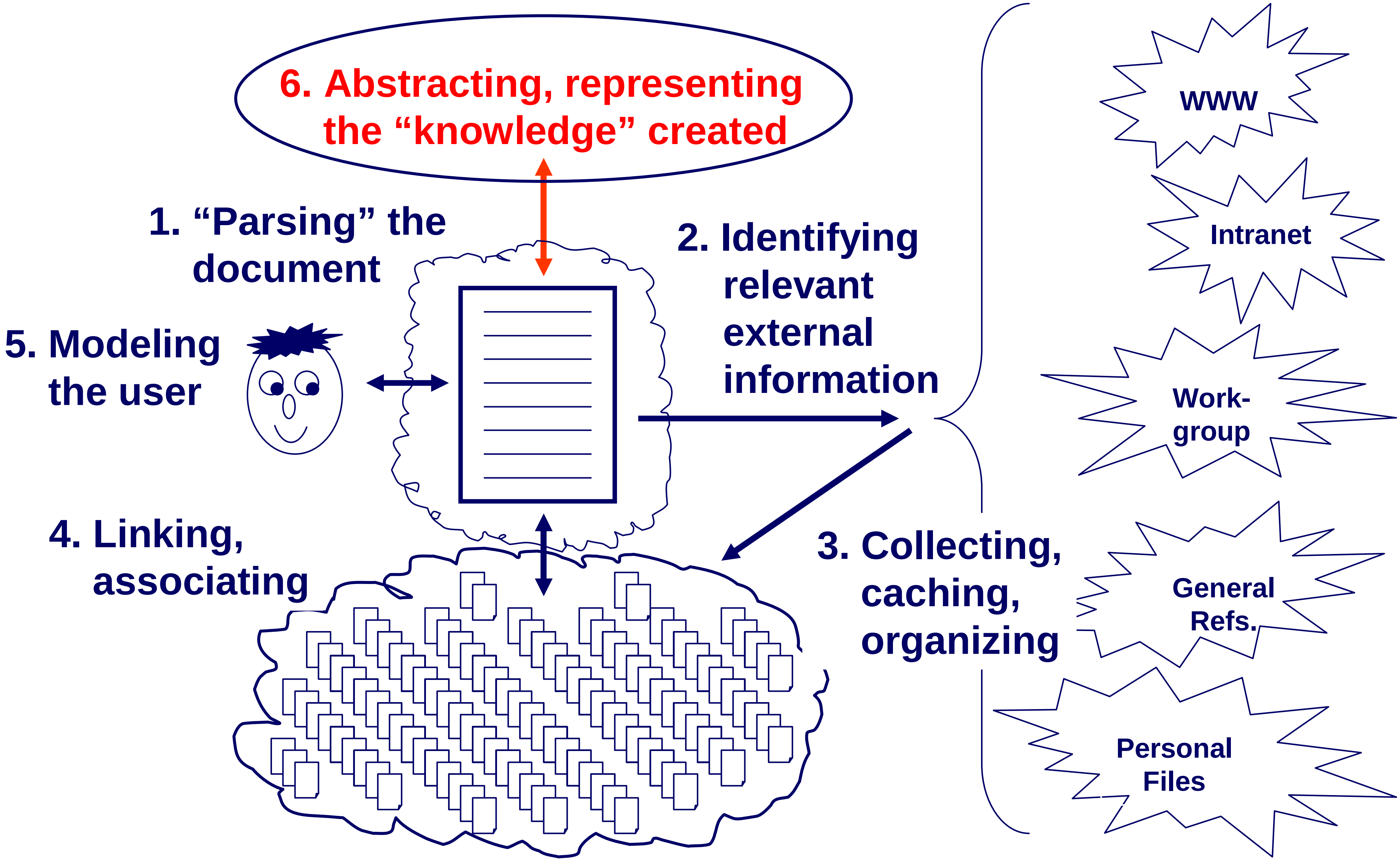
Intranet

Work-group

General Refs.

Personal Files

ViviDocs Challenges



6. Abstracting, representing the “knowledge” created

1. “Parsing” the document

2. Identifying relevant external information

5. Modeling the user

4. Linking, associating

3. Collecting, caching, organizing

WWW

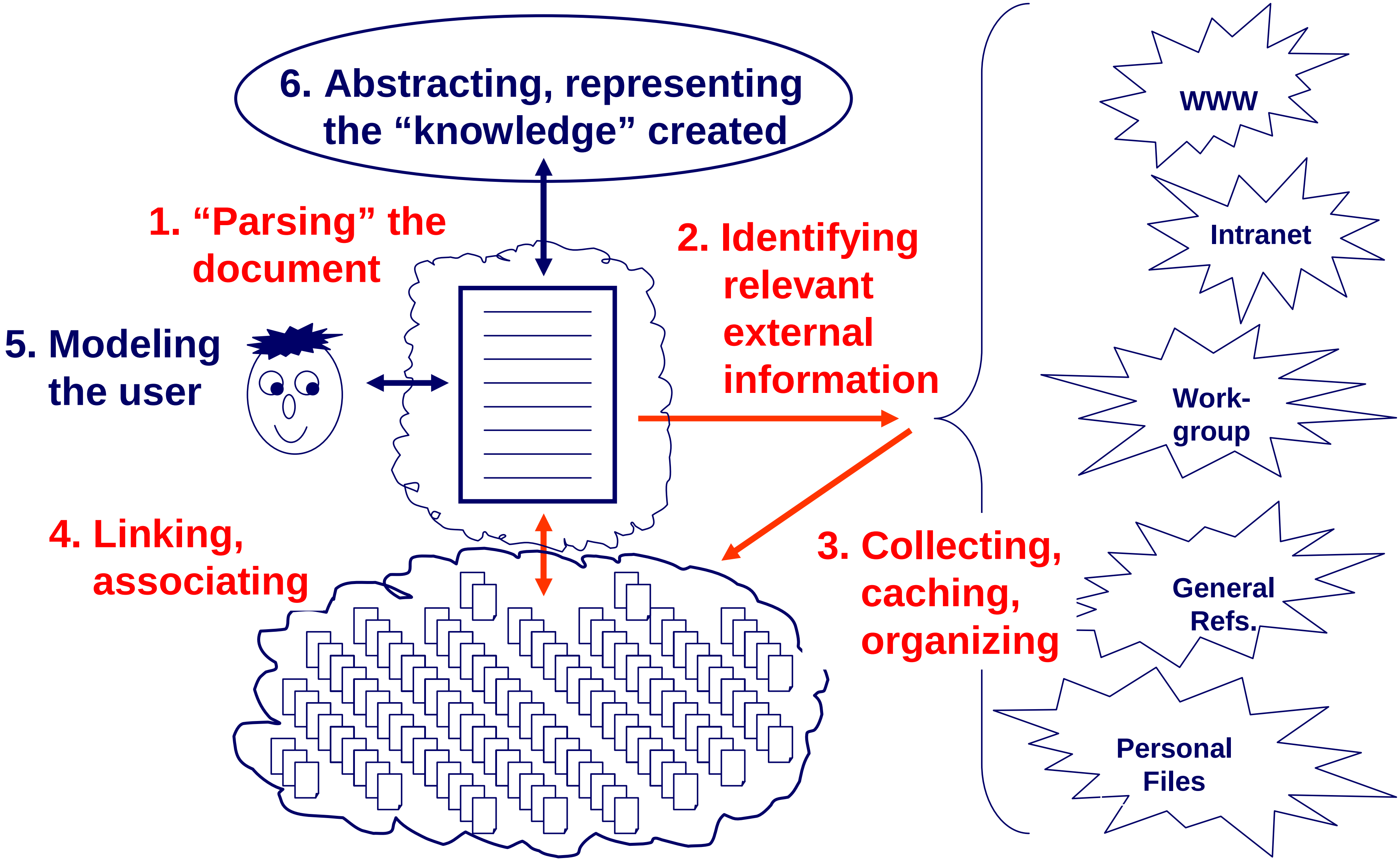
Intranet

Work-group

General Refs.

Personal Files

ViviDocs Challenges



6. Abstracting, representing the “knowledge” created

1. “Parsing” the document

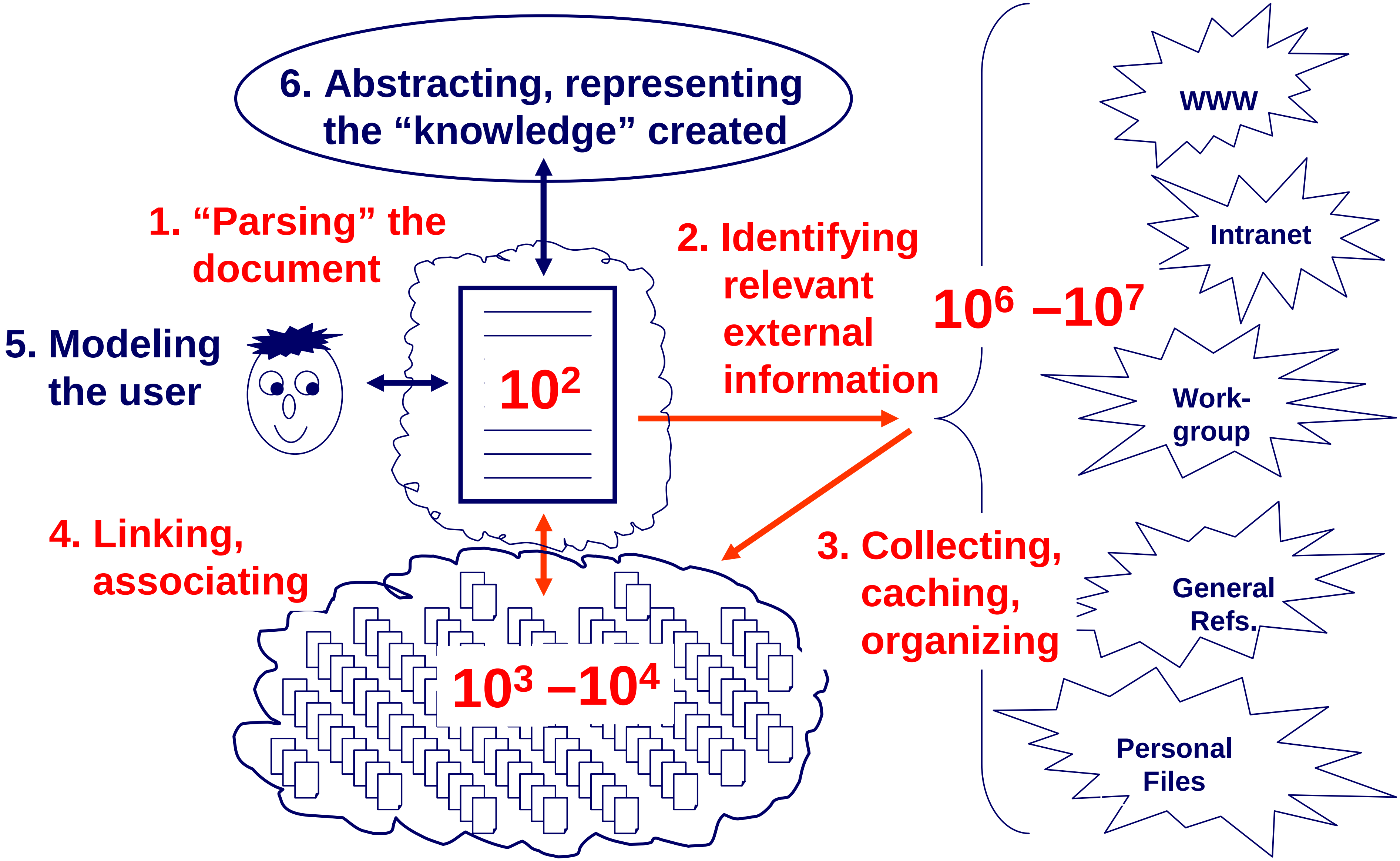
2. Identifying relevant external information

5. Modeling the user

4. Linking, associating

3. Collecting, caching, organizing

ViviDocs Challenges



6. Abstracting, representing the “knowledge” created

1. “Parsing” the document

2. Identifying relevant external information

3. Collecting, caching, organizing

4. Linking, associating

5. Modeling the user

Information Objects

- **Information Object (IO)**
  - word, entity, term, concept, phrase, proposition, sentence, paragraph, section, document, collection, …

The Battle of Stalingrad represented a major turning point for the Germany Army. The German general Paulus was out-foxed by the Russian Generals by being drawn into the city. The Russians eventually wore the Germans down, cut off their supply lines, and made retreat impossible.

Information Objects

- **Information Object (IO)**
  - word, entity, term, concept, phrase, proposition, sentence, paragraph, section, document, collection, …

The Battle of Stalingrad represented a major turning point for the Germany Army. The German general Paulus was out-foxed by the Russian Generals by being drawn into the city. The Russians eventually wore the Germans down, cut off their supply lines, and made retreat impossible.

Information Objects

- **Information Object (IO)**
  - word, entity, term, concept, phrase, proposition, sentence, paragraph, section, document, collection, …

The Battle of Stalingrad represented a major turning point for the Germany Army. The German general Paulus was out-foxed by the Russian Generals by being drawn into the city. The Russians eventually wore the Germans down, cut off their supply lines, and made retreat impossible.

Information Objects

- **Information Object (IO)**
  - word, entity, term, concept, phrase, proposition, sentence, paragraph, section, document, collection, …

The Battle of Stalingrad represented a major turning point for the Germany Army. The German general Paulus was out-foxed by the Russian Generals by being drawn into the city. The Russians eventually wore the Germans down, cut off their supply lines, and made retreat impossible.

Information Objects

- **Information Object (IO)**
  - word, entity, term, concept, phrase, proposition, sentence, paragraph, section, document, collection, …

The Battle of Stalingrad represented a major turning point for the Germany Army. The German general Paulus was out-foxed by the Russian Generals by being drawn into the city. The Russians eventually wore the Germans down, cut off their supply lines, and made retreat impossible.

Agentized, Contextualized Filters

- Independent agents attached to IOs
- Triggered by document action
- Identify an appropriate context for the IO (typically local)
- Actively fetch relevant information
- Filter relevant information
- Cache & organize information until summoned
- Establish a link / relation between IOs

Overview

- Document-Centric Systems
- ViviDocs, IOs, ACFs
- Model
- Operational Screenshots
- Conclusions

ViviDocs Agent

$$ACF_i(P_i, R_i, S_i, H_i, \theta_i, U_i, C_i, T_i, F_i)$$

P_i – represents the feature profile of information object
R_i – associated knowledge resources
S_i – target sources
H_i – history lists
θ_i – threshold
U_i – utility function for the use
C_i – processing context
T_i – triggering condition that activates the agent
F_i – response function

Necessary*, possibly *Sufficient

ViviDocs Agent

Profile

$$ACF_i(\boxed{P_i}, R_i, S_i, H_i, \theta_i, U_i, C_i, T_i, F_i)$$

The Profile is a representation of the information object based on its textual content.

For example, in an information retrieval system, a profile representing an IO (e.g., a document or paragraph) might consist of a list of terms with associated weights to reflect their importance in the document or with respect to a document collection.

ViviDocs Agent

Resource

$$ACF_i(P_i,\boxed{R_i},S_i,H_i,\theta_i,U_i,C_i,T_i,F_i)$$

Resource refers to language resources

(e.g., stop words, grammar, lexicons, etc.), knowledge resources (e.g., abstract lexical-semantic types, taxonomies or classification schemata, semantic networks, inference rules, etc.), and statistical models (e.g., term frequency and distribution counts, language models, etc.) used for processing.

ViviDocs Agent

Source

$$ACF_i(P_i,R_i,\boxed{S_i},H_i,\theta_i,U_i,C_i,T_i,F_i)$$

Source refers to the target or available information sources, accessible to the user or to the agent, in which responses to information needs may be found.

In a workgroup, this might include all the user’s files and the accessible files of the members of the user’s team or department. In a business setting, this might include the intranet, extranet, and, selectively, the internet, as well as the user’s personal files.

ViviDocs Agent

History

$$ACF_i(P_i,R_i,S_i,\boxed{H_i},\theta_i,U_i,C_i,T_i,F_i)$$

History consists of lists of IOs (and perhaps “scores”) that have been generated by previous actions of ACFs.

For example, in information retrieval with user feedback, the initial ranked list of documents considered as relevant by the system can be regarded as the history for the next round of retrieval with additional user feedback.

ViviDocs Agent

Threshold

$$ACF_i(P_i,R_i,S_i,H_i,\boxed{\theta_i},U_i,C_i,T_i,F_i)$$

Threshold establishes (and controls) the cut-off point in selecting (ranking, associating, etc.) information.

Thresholds can be absolute numbers (e.g., the top 100 documents or passages), similarity scores, or confidence scores applied to retrieved information.

ViviDocs Agent

Utility

$$ACF_i(P_i, R_i, S_i, H_i, \theta_i, \boxed{U_i}, C_i, T_i, F_i)$$

Utility is used to measure and rank system outputs based on their benefits for the user or on the degree to which they satisfy the user’s information needs minus the associated costs.

Such measures are commonly used in information filtering and typically calculated from an explicit or implicit tolerance for “noise” (the ratio of true-positive to false-positive responses) in the output.

ViviDocs Agent

Context

$$ACF_i(P_i,R_i,S_i,H_i,\theta_i,U_i,\boxed{C_i},T_i,F_i)$$

Context provides additional information that can be associated with the profile.

This concept is inherently open-ended; we restrict it to information that is operationally available to the system. We distinguish at least three kinds of context: (a) global context, (b) local context, and (c) focus. In an IR-like action anchored to a specific IO (e.g., word or phrase), the global context might be the document in which the IO occurs; the local context the paragraph; the focus the sentence (essentially, the proposition expressed).

ViviDocs Agent

$$ACF_i(P_i,R_i,S_i,H_i,\theta_i,U_i,C_i,\overset{\text{Trigger}}{\boxed{T_i}},F_i)$$

Triggers activate the ACFs.

The action associated with opening a document or beginning to compose a message could launch a battery of ACFs. Under a GUI, triggers can take the form of highlighting, typing, clicking, etc. For example, every time the user types a full stop, an ACF can be triggered on the most recently completed sentence. Likewise ACFs could be triggered every twenty-four hours, updating the information that they associate with the IOs they are attached to.

ViviDocs Agent

Function

$$ACF_i(P_i, R_i, S_i, H_i, \theta_i, U_i, C_i, T_i, \boxed{F_i})$$

Function specifies the relation that is to be established between the IO and other information by the ACF, including the format for extracting or presenting such information.

The function might be as simple as “retrieval”—finding a rank-ordered list of documents or passages—or “answer” (a simple sentence) in response to an implicit question. But the function might be more complex…

Schematic ACF

FindRelevantDocs

FindRelevantDocs

Profile: <terms in *Passage*$_i$ ∈ *Document*, passage-count=*l*>

Resource: <<NLP Lexicon>, <NLP Grammar>, <Reference Stats>>

Source: <specified *Source*>

History: <empty>

Threshold: <all documents *d* in *Source* to rank =
max(*n*,min(count(norm-score(*d*)≥0.7),*m*)),
where n=100/*l* and *m*=10,000/*l*>

Utility: <not defined>

Context: <empty>

Trigger: <opening of *Document*>

Function: <retrieve documents from *Source* for each *Passage*$_i$; cache results>

Writing Example

“Hostage taking is a contemporary crisis…”

Writing Example

Representing Content

"Hostage taking is a contemporary crisis…"

hostage taking
hostage
taking
contemporary crisis
contemporary
crisis

Set of terms

Writing Example

Representing Content + Local Context

"The storming of the U.S. embassy in Tehran in 1979 was merely a prelude of hostilities to come. Hostage taking is a contemporary crisis…"

storming
u.s. embassy
u.s.
embassy
tehran
1979
prelude
hostilities

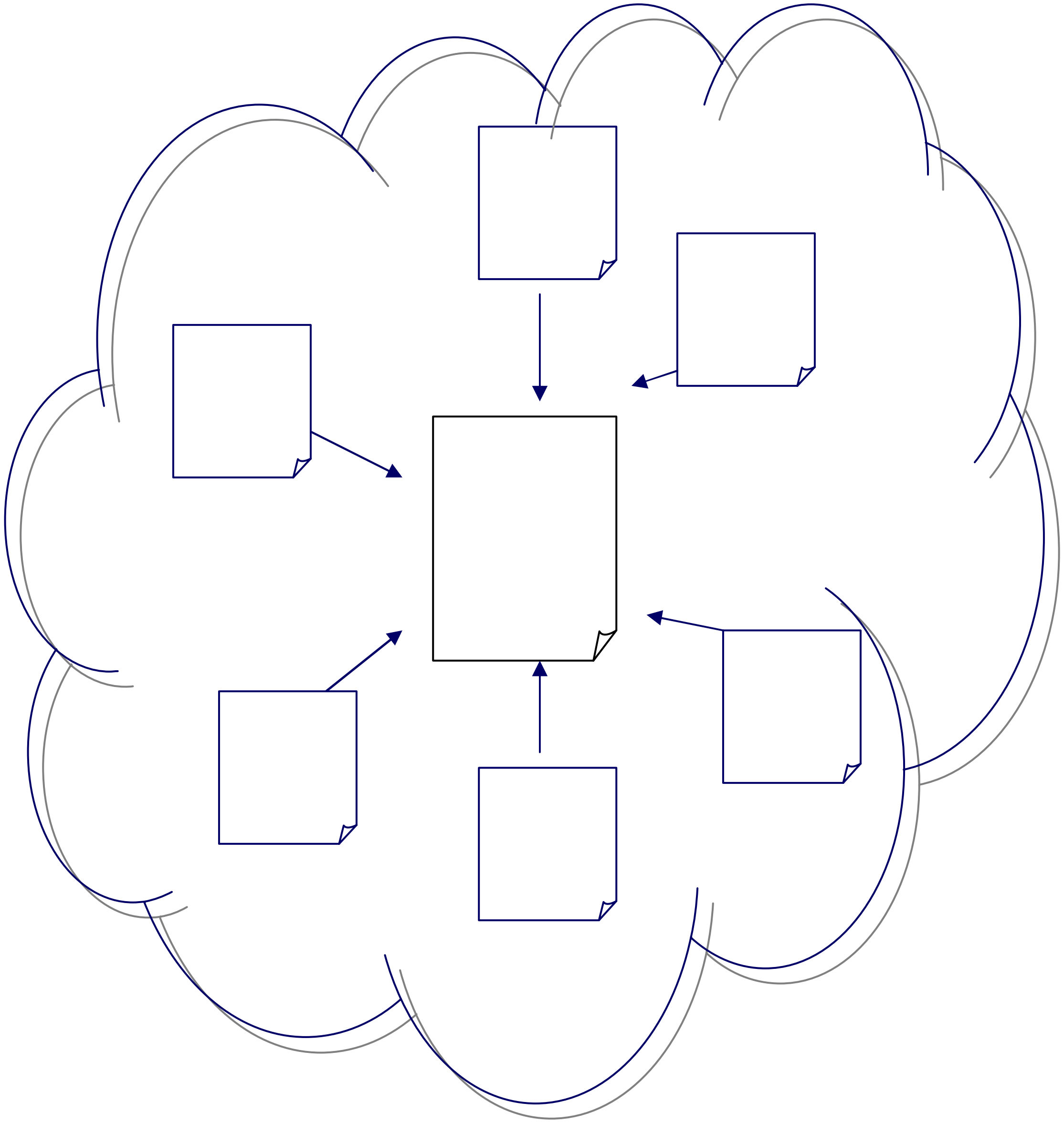
+

hostage taking
hostage
taking
contemporary crisis
contemporary
crisis

Composite Term Vector

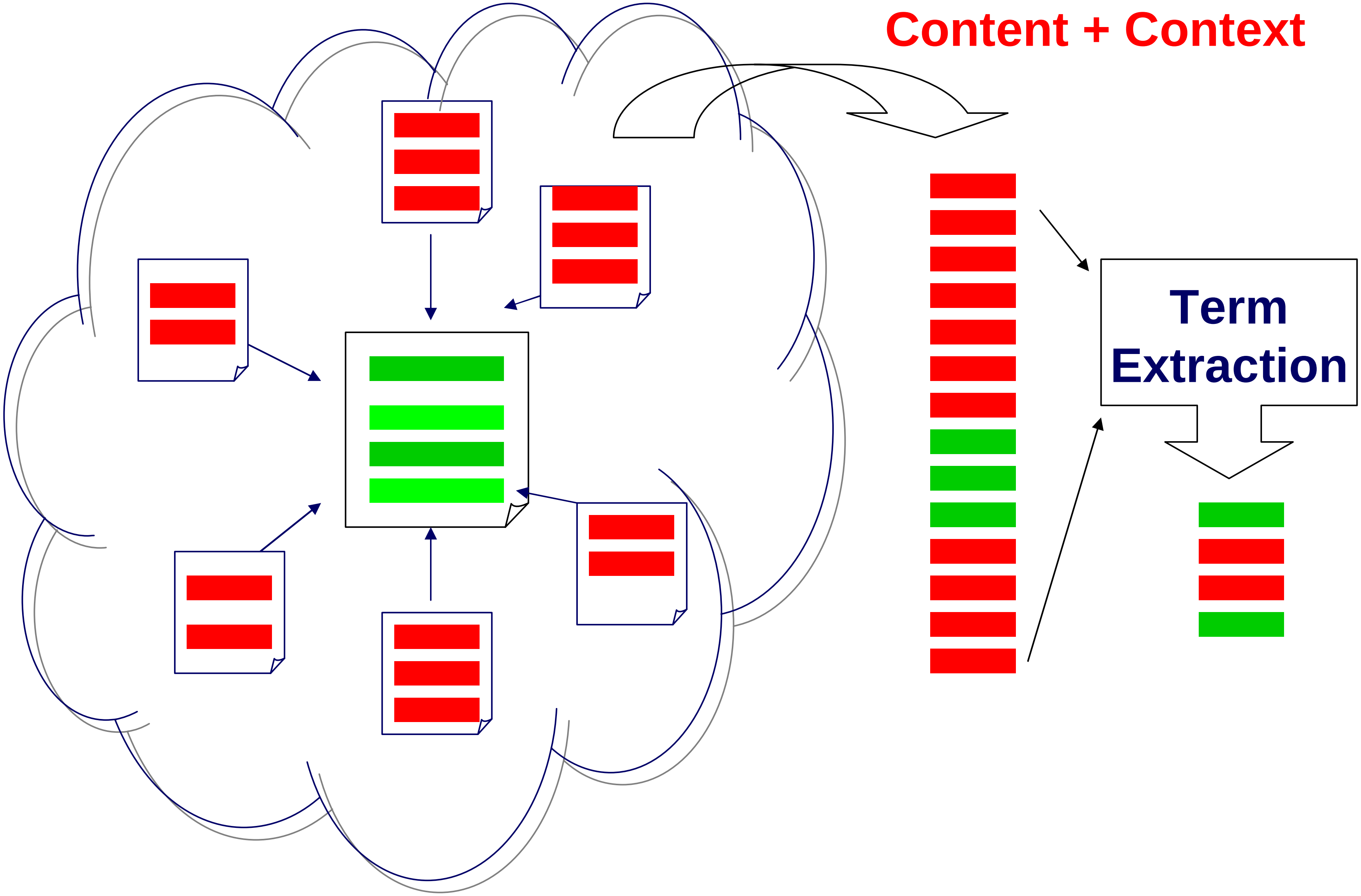
1.00 (hostage taking)
1.00 (hostage)
1.00 (taking)
1.00 (contemporary crisis)
1.00 (contemporary)
1.00 (crisis)
0.25 (storming)
0.25 (u.s. embassy)
0.25 (u.s.)
0.25 (embassy)
0.25 (tehran)
0.25 (1979)
0.25 (prelude)
0.25 (hostilities)

Document in Context of Work

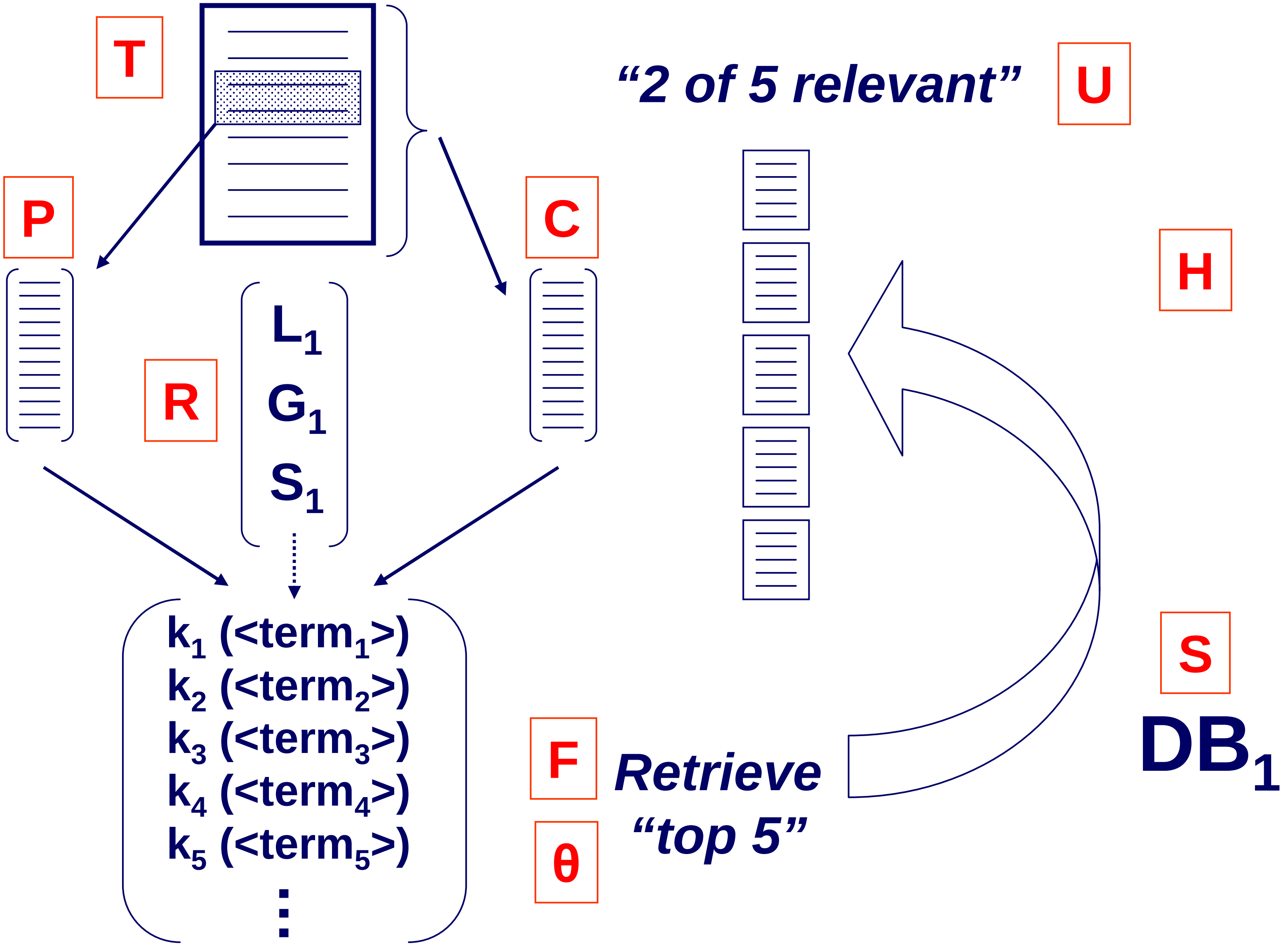


Document in Context of Work

Content + Context



Schematic Illustration of ACF Action



T

P

C

R

L_1

G_1

S_1

k_1 (<term_1>)

k_2 (<term_2>)

k_3 (<term_3>)

k_4 (<term_4>)

k_5 (<term_5>)

⋮

F

θ

Retrieve “top 5”

“2 of 5 relevant”

U

H

S

DB_1

Instantiated ACF

FindRelevantDocs

FindRelevantDocs

Profile:	**<*contemporary crisis*: 0;**
	***hostage taking*: 22;**
	***hostage*: 587;**
	***contemporary*: 2387;**
	***crisis*: 4149;**
	***taking*: 12042 >**

Resource:	**<English Lexicon, English Grammar, AP88 DB Stats>**
Source:	**<indexed AP88 DB with 3-sentence passages>**
History:	**<empty>**
Threshold:	**<N=100>**
Utility:	**<not defined>**
Context:	**<empty>**
Trigger:	**<typing of “.”>**
Function:	**<retrieval ; caching (=IO_2)>**

Instantiated ACF

FindDescriptionWhere

<u>FindDescriptionWhere</u>

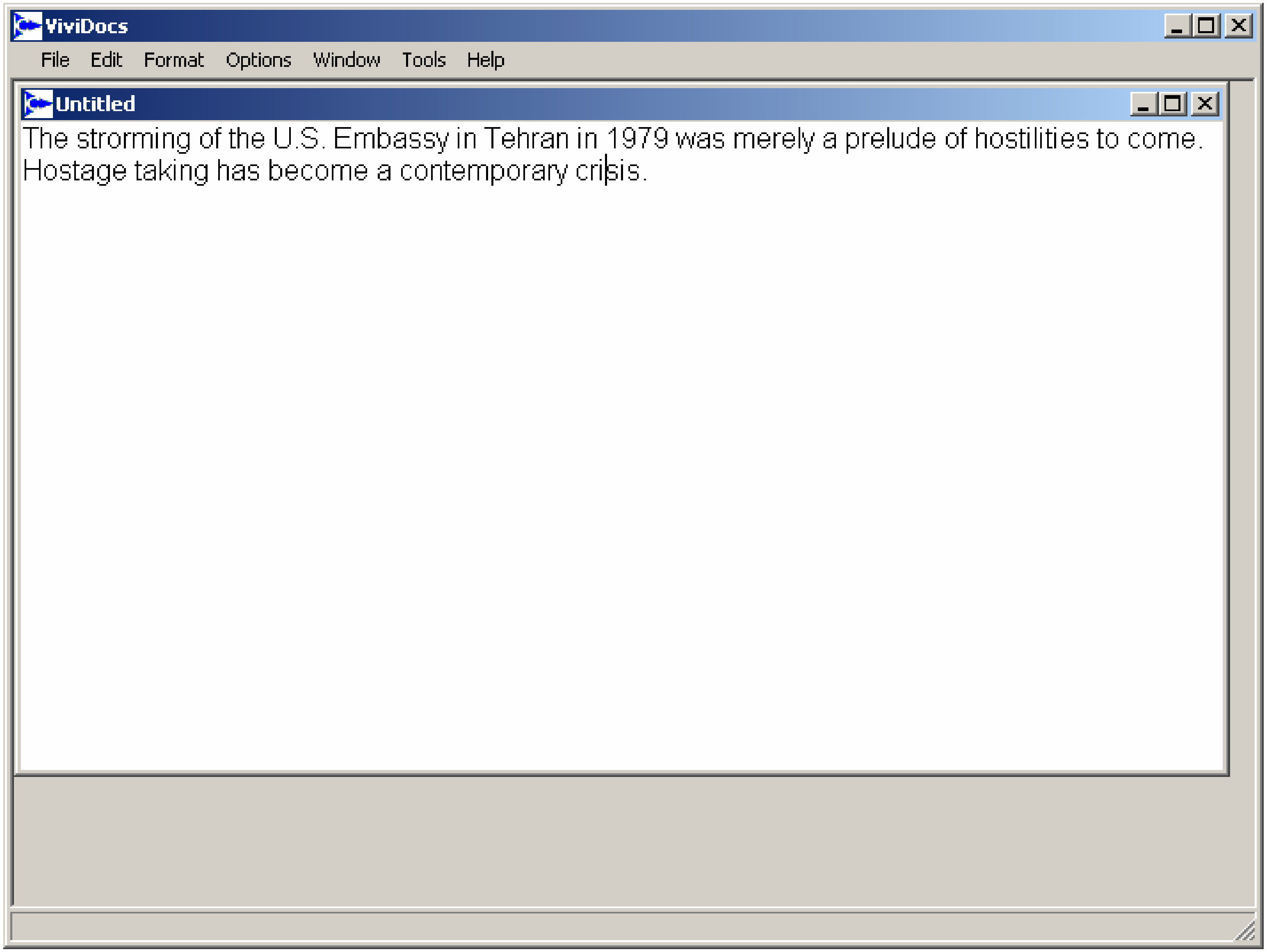
Profile:	**<*contemporary crisis*: 0;
	hostage taking: 22;
	hostage: 587;
	contemporary: 2387;
	crisis: 4149;
	taking: 12042 >**

Resource:	**<English Lexicon, English Grammar, AP88 DB Stats>**
Source:	**<indexed database built based on IO_2>**
History:	**<IO_2>**
Threshold:	**<N=10>**
Utility:	**<not defined>**
Context:	**<IO_2>**
Trigger:	**<mouse click and menu selection>**
Function:	**<answer-where>**

Overview

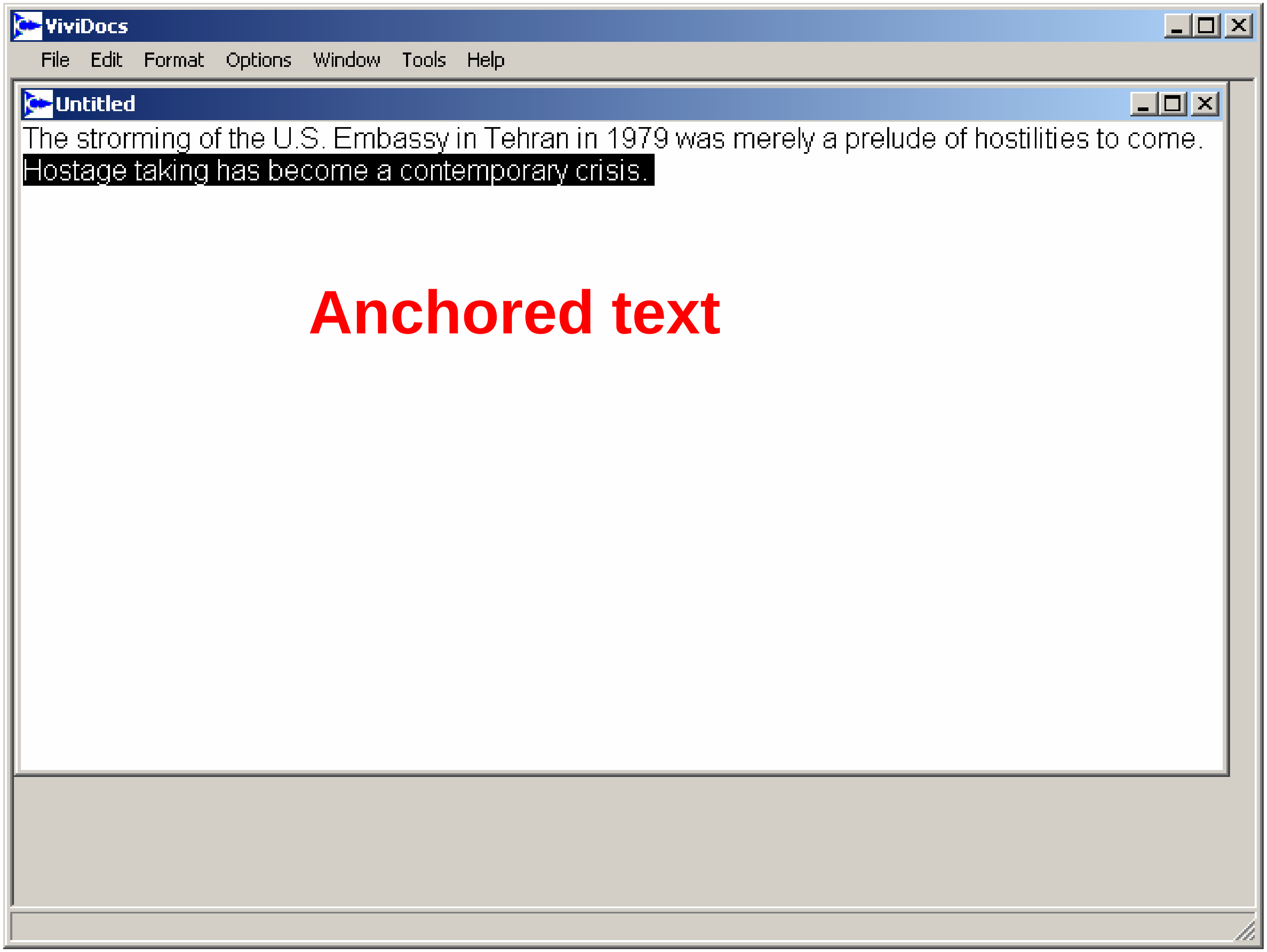
- Document-Centric Systems
- ViviDocs, IOs, ACFs
- Model
- Operational Screenshots
- Conclusions

ViviDocs Example



The stroming of the U.S. Embassy in Tehran in 1979 was merely a prelude of hostilities to come. Hostage taking has become a contemporary crisis.

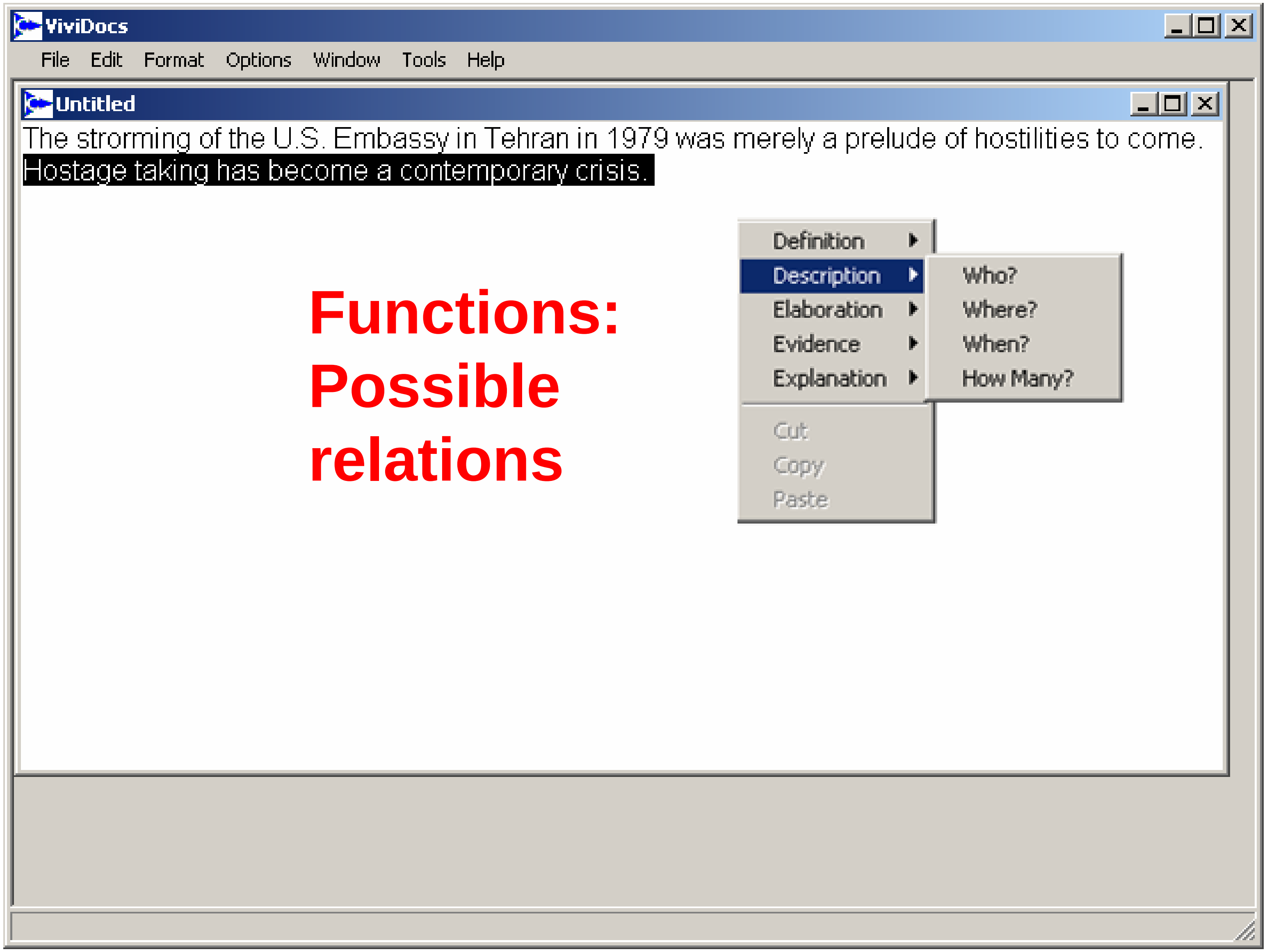
ViviDocs Example



The strorming of the U.S. Embassy in Tehran in 1979 was merely a prelude of hostilities to come. Hostage taking has become a contemporary crisis.

Anchored text

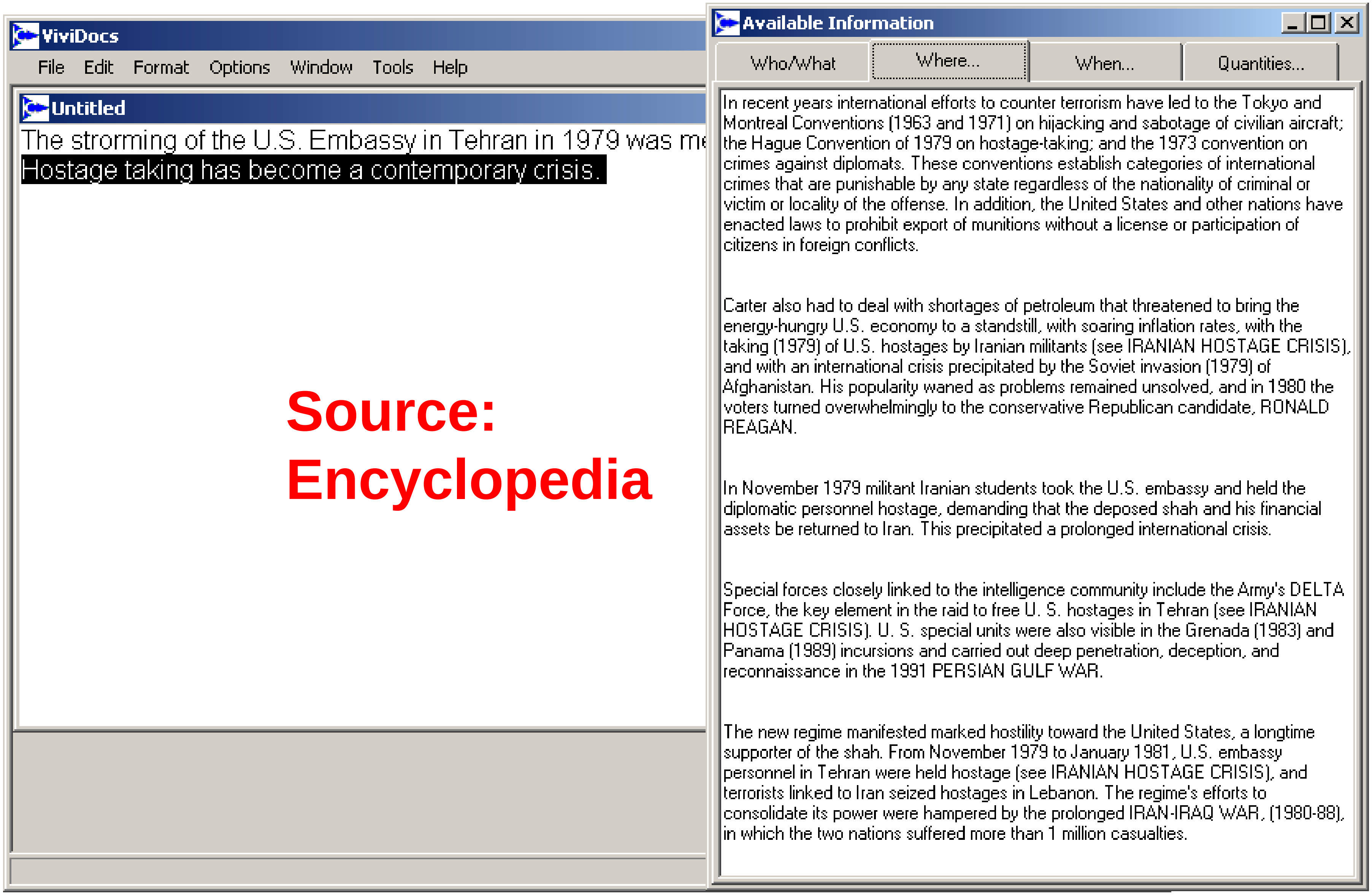
ViviDocs Example



The strorming of the U.S. Embassy in Tehran in 1979 was merely a prelude of hostilities to come. Hostage taking has become a contemporary crisis.

Functions: Possible relations

ViviDocs Example



The strorming of the U.S. Embassy in Tehran in 1979 was me

Hostage taking has become a contemporary crisis.

Source: Encyclopedia

In recent years international efforts to counter terrorism have led to the Tokyo and Montreal Conventions (1963 and 1971) on hijacking and sabotage of civilian aircraft; the Hague Convention of 1979 on hostage-taking; and the 1973 convention on crimes against diplomats. These conventions establish categories of international crimes that are punishable by any state regardless of the nationality of criminal or victim or locality of the offense. In addition, the United States and other nations have enacted laws to prohibit export of munitions without a license or participation of citizens in foreign conflicts.

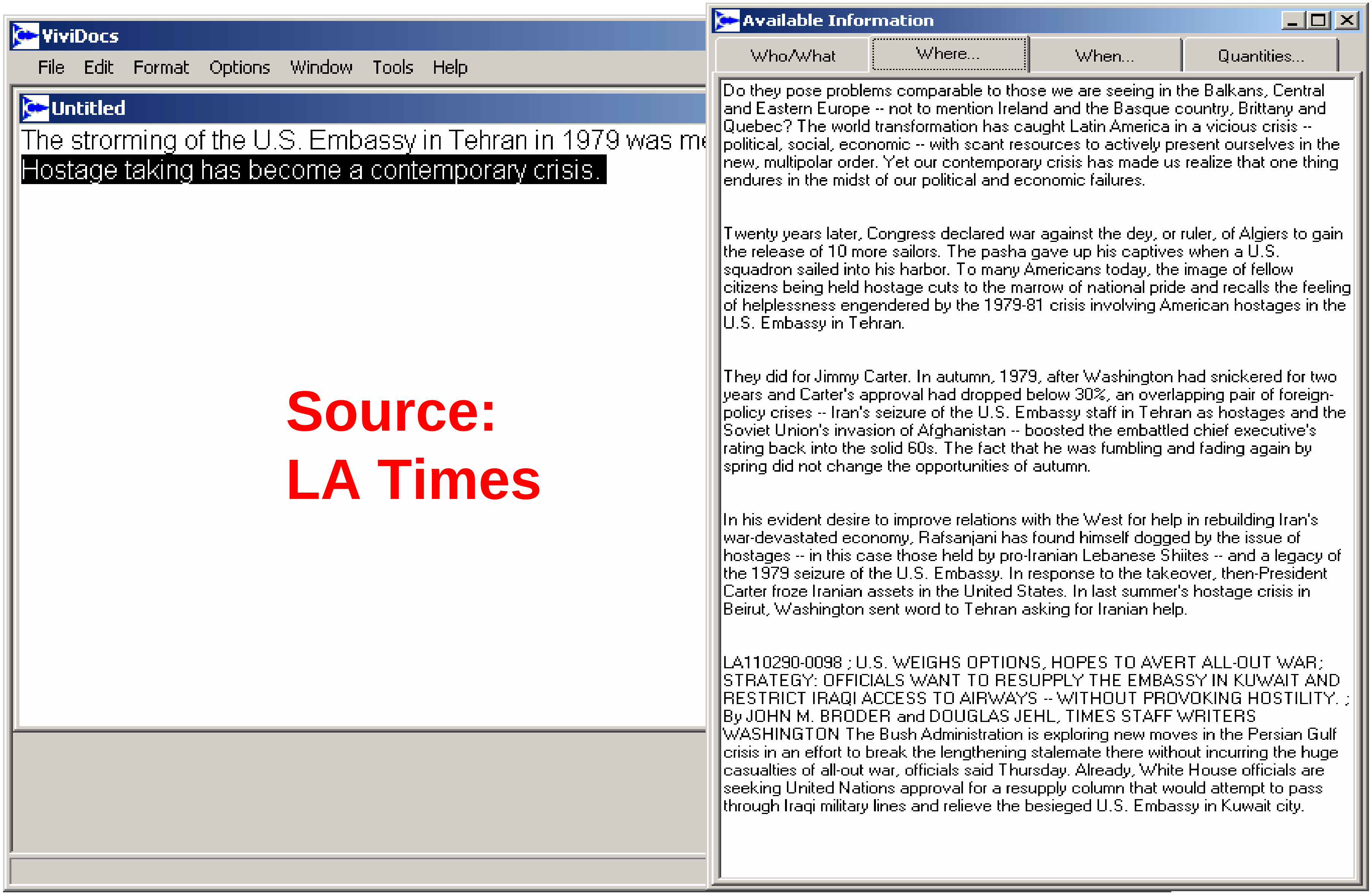
Carter also had to deal with shortages of petroleum that threatened to bring the energy-hungry U.S. economy to a standstill, with soaring inflation rates, with the taking (1979) of U.S. hostages by Iranian militants (see IRANIAN HOSTAGE CRISIS), and with an international crisis precipitated by the Soviet invasion (1979) of Afghanistan. His popularity waned as problems remained unsolved, and in 1980 the voters turned overwhelmingly to the conservative Republican candidate, RONALD REAGAN.

In November 1979 militant Iranian students took the U.S. embassy and held the diplomatic personnel hostage, demanding that the deposed shah and his financial assets be returned to Iran. This precipitated a prolonged international crisis.

Special forces closely linked to the intelligence community include the Army's DELTA Force, the key element in the raid to free U. S. hostages in Tehran (see IRANIAN HOSTAGE CRISIS). U. S. special units were also visible in the Grenada (1983) and Panama (1989) incursions and carried out deep penetration, deception, and reconnaissance in the 1991 PERSIAN GULF WAR.

The new regime manifested marked hostility toward the United States, a longtime supporter of the shah. From November 1979 to January 1981, U.S. embassy personnel in Tehran were held hostage (see IRANIAN HOSTAGE CRISIS), and terrorists linked to Iran seized hostages in Lebanon. The regime's efforts to consolidate its power were hampered by the prolonged IRAN-IRAQ WAR, (1980-88), in which the two nations suffered more than 1 million casualties.

ViviDocs Example



The strorming of the U.S. Embassy in Tehran in 1979 was me

Hostage taking has become a contemporary crisis.

Do they pose problems comparable to those we are seeing in the Balkans, Central and Eastern Europe -- not to mention Ireland and the Basque country, Brittany and Quebec? The world transformation has caught Latin America in a vicious crisis -- political, social, economic -- with scant resources to actively present ourselves in the new, multipolar order. Yet our contemporary crisis has made us realize that one thing endures in the midst of our political and economic failures.

Twenty years later, Congress declared war against the dey, or ruler, of Algiers to gain the release of 10 more sailors. The pasha gave up his captives when a U.S. squadron sailed into his harbor. To many Americans today, the image of fellow citizens being held hostage cuts to the marrow of national pride and recalls the feeling of helplessness engendered by the 1979-81 crisis involving American hostages in the U.S. Embassy in Tehran.

They did for Jimmy Carter. In autumn, 1979, after Washington had snickered for two years and Carter's approval had dropped below 30%, an overlapping pair of foreign-policy crises -- Iran's seizure of the U.S. Embassy staff in Tehran as hostages and the Soviet Union's invasion of Afghanistan -- boosted the embattled chief executive's rating back into the solid 60s. The fact that he was fumbling and fading again by spring did not change the opportunities of autumn.

In his evident desire to improve relations with the West for help in rebuilding Iran's war-devastated economy, Rafsanjani has found himself dogged by the issue of hostages -- in this case those held by pro-Iranian Lebanese Shiites -- and a legacy of the 1979 seizure of the U.S. Embassy. In response to the takeover, then-President Carter froze Iranian assets in the United States. In last summer's hostage crisis in Beirut, Washington sent word to Tehran asking for Iranian help.

LA110290-0098 ; U.S. WEIGHS OPTIONS, HOPES TO AVERT ALL-OUT WAR; STRATEGY: OFFICIALS WANT TO RESUPPLY THE EMBASSY IN KUWAIT AND RESTRICT IRAQI ACCESS TO AIRWAYS -- WITHOUT PROVOKING HOSTILITY. ; By JOHN M. BRODER and DOUGLAS JEHL, TIMES STAFF WRITERS WASHINGTON The Bush Administration is exploring new moves in the Persian Gulf crisis in an effort to break the lengthening stalemate there without incurring the huge casualties of all-out war, officials said Thursday. Already, White House officials are seeking United Nations approval for a resupply column that would attempt to pass through Iraqi military lines and relieve the besieged U.S. Embassy in Kuwait city.

Source: LA Times

Overview

- Document-Centric Systems
- ViviDocs, IOs, ACFs
- Model
- Operational Screenshots
- Conclusions

Conclusion

- **Increasing need for document-centric processing in support of user**
- **ViviDocs is one example; early-stage prototype**
- **Necessary (but not sufficient?) operations captured by ACFs**
- **Possibly 25–50 ACF types may handle all interesting cases (6–8 explored to date)**
- **Work remains in all six “challenging” areas**
  - Decomposing the document (“parsing”)
  - Selecting (caching) appropriate data
  - Organizing (analyzing) gathered data
  - Linking information to document
  - Modeling and integrating the user
  - Representing the knowledge/information developed

The End