



Semantics and Experience in the Future Web

Enric Plaza
IIIA-CSIC

ECCBR-2008, Trier, Rheinland-Pfalz



Outline

Introduction

Semantics, Up&Down

The Network is the Content

The Case for Experience

Reusing Other People's Experiences

Semantics and Experience

Forms of Experience

The EDIR Cycle

Discussion/Challenges



Introduction

Use the web for...



Introduction

Use the web for...

Find something
(gain access to some information)

Do something (in the world)
Take a **decision** (in the world)



Introduction

Web of Documents

Large amount of
experiences of
individual people

But they are treated as
documents in blogs,
Q&A sites, forums, social
software



Introduction

documents

ount of
ces of
people

treated as
s in blogs,
ums, social
are

There is a special form of
content, **experiential
knowledge**,
that should be represented,
organized, analyzed, retrieved
and reused as **such
experiences**
(and not as documents)

Experiences are probably
the most added-value assets on
the web



Semantics, Up & Down



Semantics, Up & Down

Top-down

Bottom-up



Semantics, Up & Down

Top-down



**Semantic
Web**

from human-readable to
machine-readable, to
provide service
information exchange



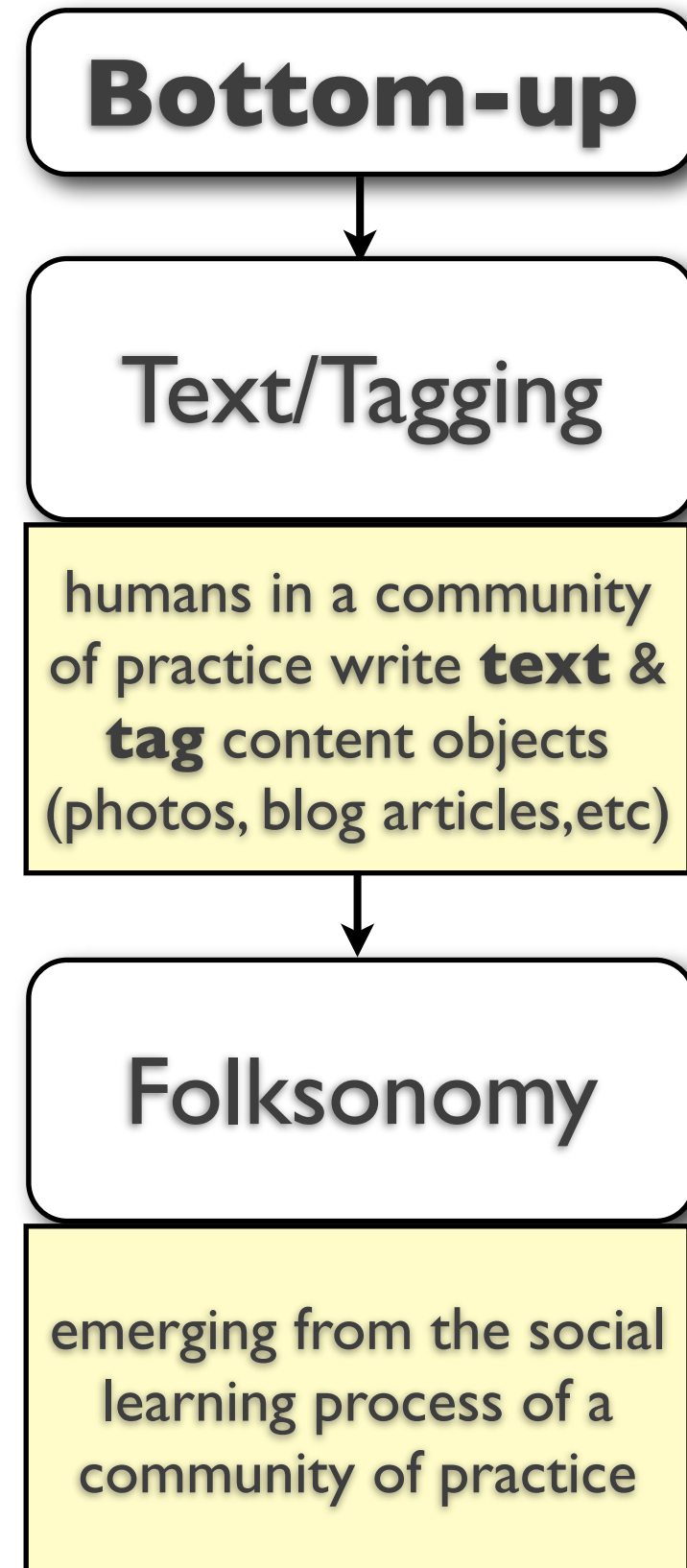
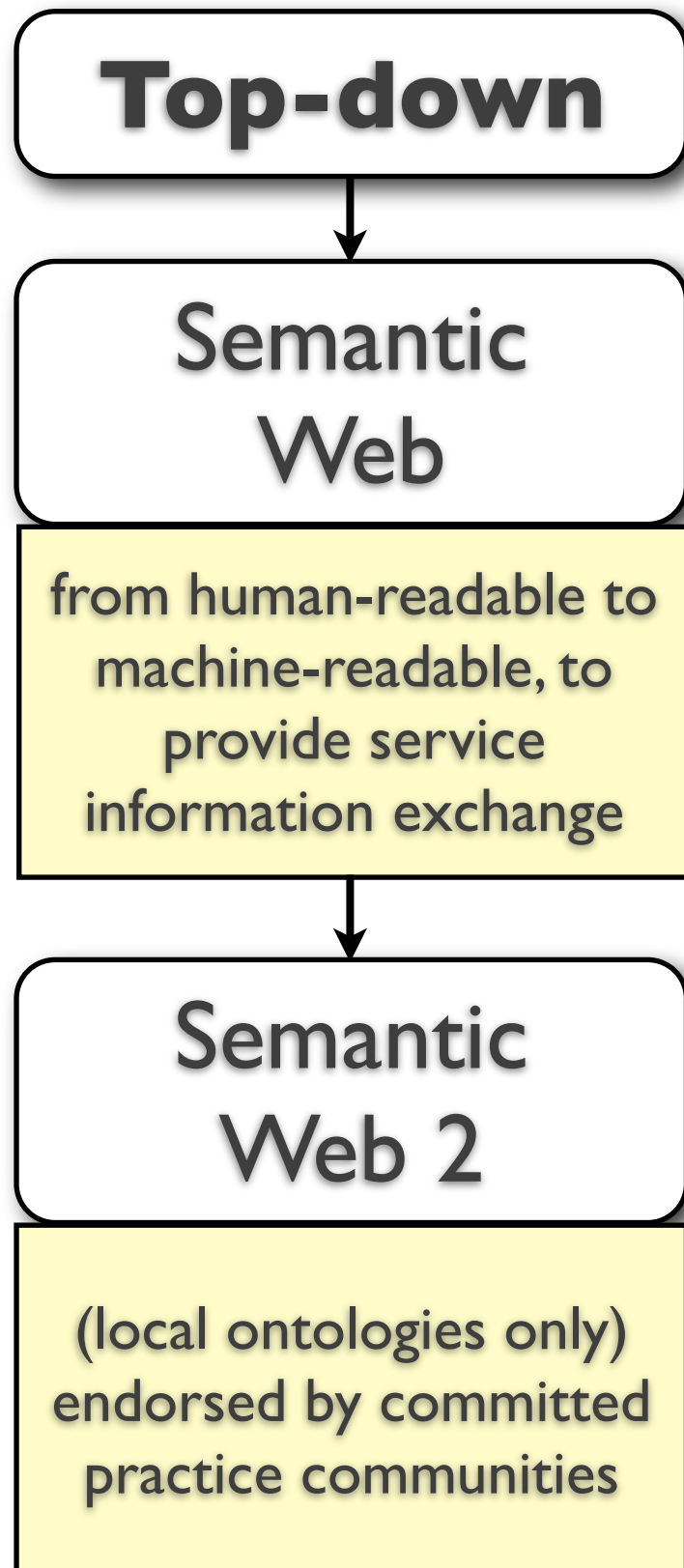
**Semantic
Web 2**

(local ontologies only)
endorsed by committed
practice communities

Bottom-up



Semantics, Up & Down





Semantics, Up & Down

Top-down

Bottom-up



Semantics, Up & Down

Top-down



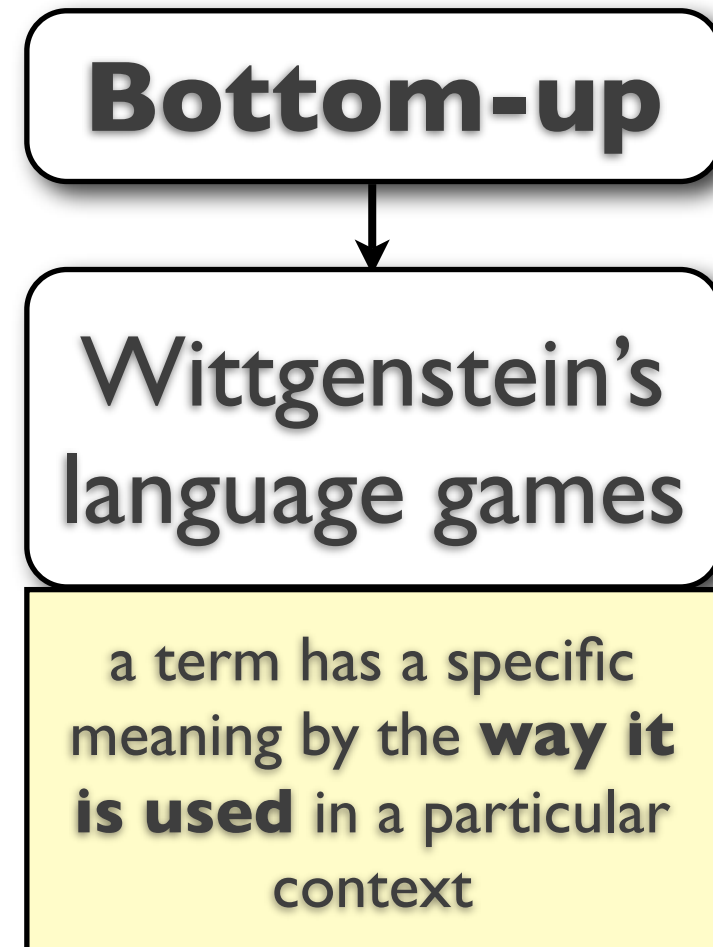
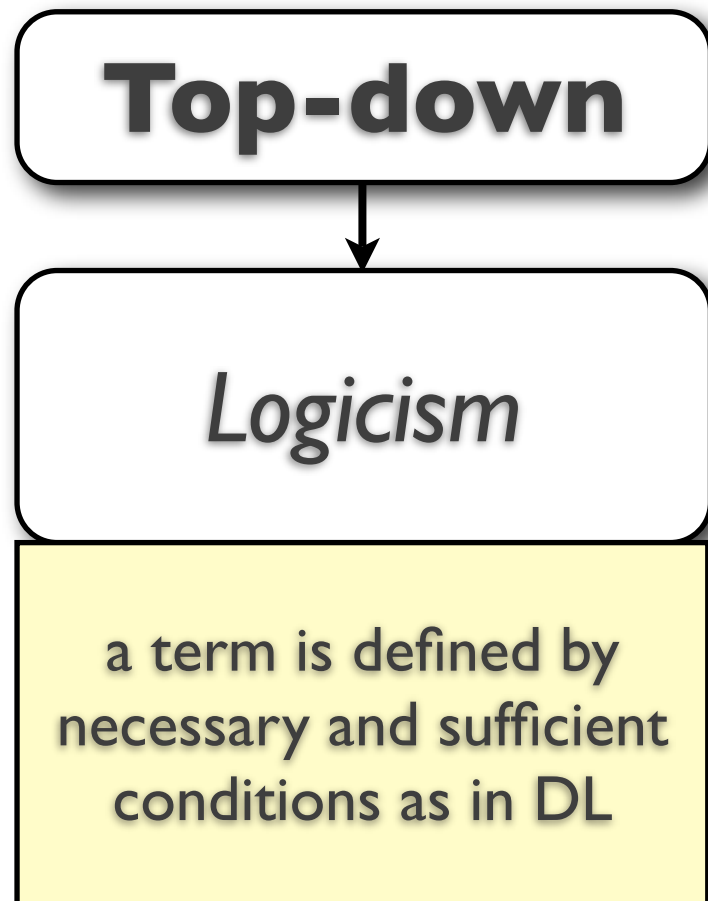
Logicism

a term is defined by
necessary and sufficient
conditions as in DL

Bottom-up

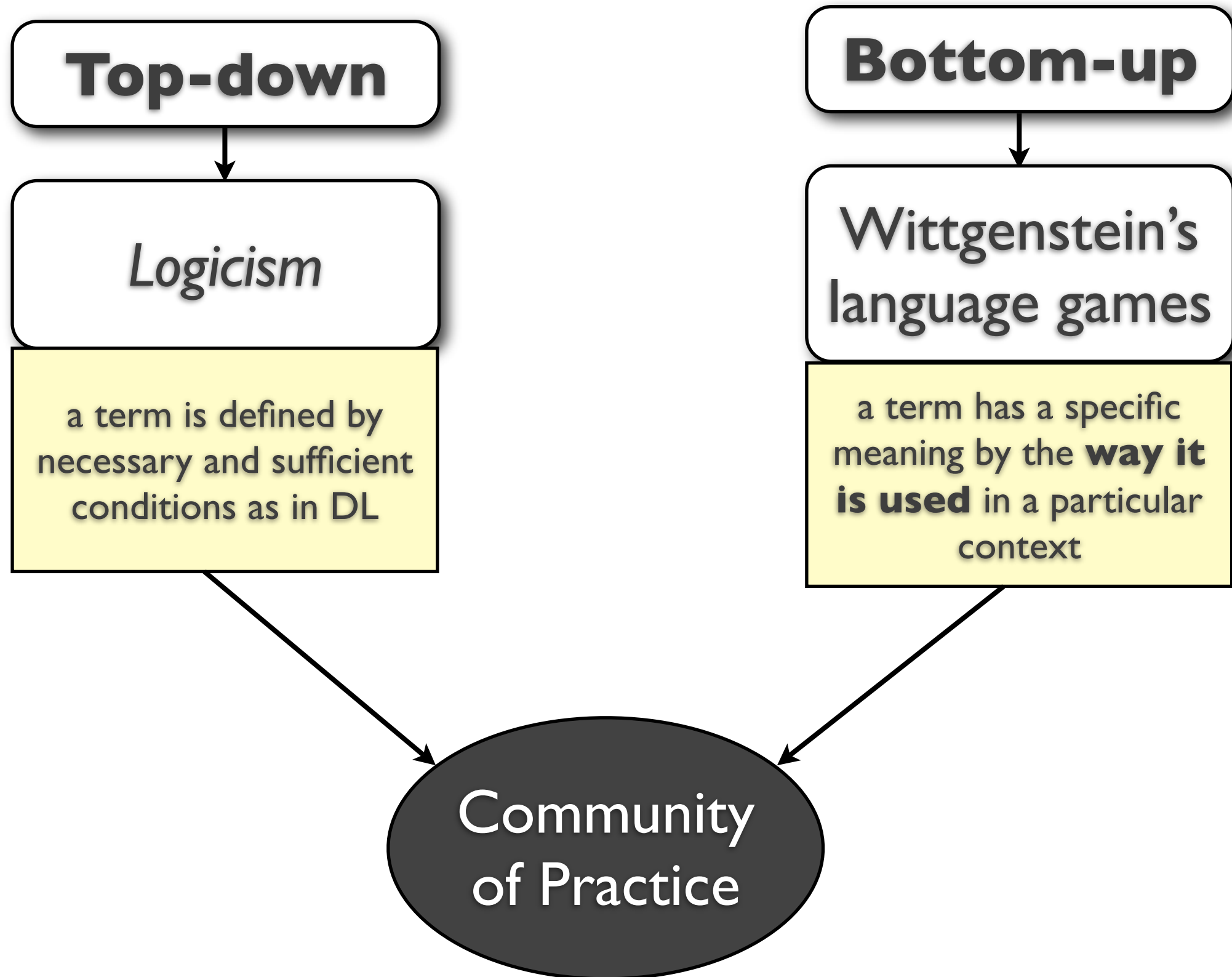


Semantics, Up & Down





Semantics, Up & Down





Semantic Enabling

Communities of Practice

Hybrid top-down & bottom-up approaches with approximate concept descriptions

Enabling Technologies

Semantic web, ontologies, folksonomies are needed as a substrate that provides some service required by more complex tasks

Empirical Issues

Which approaches are more suitable to capture explicit knowledge, tacit knowledge

Which approaches are more suitable to different forms of content

Which approaches are more suitable to different web-based systems



The Network is the Content (or vice versa)



Network effects

Web 2.0 & Social networks

Social networking among individuals is heralded as the the most important innovation; this would mean establishing **social relationships** by means of the web is what is creating new knowledge and new value

User-contributed content

Declaring new relationships is a form of user-contributed content

User-contributed relationships may be among people, but also other forms of content: e.g. tagging photos in *Flickr*

Google basically analyzed user-contributed hyperlink relationships among pages to estimate page importance/significance



The Case for Experience



Experience

CBR

Case-based reasoning may be understood as learning to make better decisions or predictions from past experience

situation₃

outcome₃



Experience

CBR

Case-based reasoning may be understood as learning to make better decisions or predictions from past experience

situation₃

outcome₃

Experience:

knowledge about an observed factual situation

“This is a good hotel because my stay was very agreeable”

“I did this sequence of actions, in this situation, and I achieved this goal”

Although there are no “explicit (s₃,o₃) cases” on the web, there is a huge amount of **practical knowledge** present on the web; this kind of practical knowledge coming from direct observations (experiences) is what we’ll call **experience**



Experiential Content

Most valuable asset

People constantly search & browse the web resources to find other peoples experiences in a solving given problem, achieving a particular goal, obtaining a particular outcome, or deciding some issue



Experiential Content

Most valuable asset

People constantly search & browse the web resources to find other peoples experiences in a solving given problem, achieving a particular goal, obtaining a particular outcome, or deciding some issue

Hypothesis: there is “experiential content”

People use it to decide which hotel to book, which spots to visit

People Browse websites/forums on digital photography to learn how to solve issues they encountered with their photos

The challenge is how to represent, organize, and reuse experiential content beyond a collection of hyperlinked documents



Search & Browse

Unsupported user's task in S&B

People search and browse in the same unsupported way, independently of whether they are googling the Web, or searching in a thematic website (e.g. forums)

Search: yields a large amount of “resources”

Browse: user has to read large collections of “found items” and find what is interesting for her **purpose**

Filter: eliminating irrelevant found items is unsupported; usually just copy & paste interesting items

Reuse: user analyzes the content of the relevant retrieved items and takes a decision according to her purpose



Found & Lost

H: hotels in an intended destination

W: websites with hotel-related experiential content destination

C: average number of client reports per hotel

S&B: $H \times W \times C$ user-contributed experience items
[Impossible to be manually processed by the user]



Found & Lost

H: hotels in an intended destination

W: websites with hotel-related experiential content destination

C: average number of client reports per hotel

S&B: $H \times W \times C$ user-contributed experience items
[Impossible to be manually processed by the user]

h=filter(H): 3-star hotels only

w=sample(W): visit only a few websites destination

c=sample(C): read only a few reports

S&B: $h \times w \times c$

However there is no computer support to obtaining **good samples** of performing **good filters**



Reuse

The real task

- A) Need to aggregate for each hotel **h** pros and cons according to the majority opinion of the **w** x **c** reports
- B) Finally decide which hotel fits better my purposes (one-night business trip vs. family week vacation trip)



Reuse

The real task

- A) Need to aggregate for each hotel **h** pros and cons according to the majority opinion of the **w** x **c** reports
- B) Finally decide which hotel fits better my purposes (one-night business trip vs. family week vacation trip)

An alternative approach that support users
in making more informed decision

A reinterpretation of CBR that supporting the reuse of
experiential knowledge provided by other people
but integral to a community of practice



Reusing Other People's Experiences



S&B as Retrieve & Reuse

1) **Retrieve** process searches for client reports of hotels close to the **interests** of the user and selects a subset of them

2) **Reuse** process analyzes them in order to **aggregate** information about pros and cons of each hotel and produces a **ranking** of hotels taking into account the users interests and the pros and cons of each hotel



S&B as Retrieve & Reuse

1) **Retrieve** process searches for client reports of hotels close to the **interests** of the user and selects a subset of them

2) **Reuse** process analyzes them in order to **aggregate** information about pros and cons of each hotel and produces a **ranking** of hotels taking into account the users interests and the pros and cons of each hotel



S&B as Retrieve & Reuse

1) **Retrieve** process searches for client reports of hotels close to the **interests** of the user and selects a subset of them

*1) Given a problem (a specific task to be achieved) the **Retrieve** process selects the subset of cases (experiential knowledge) most similar (relevant) to that problem*

2) **Reuse** process analyzes them in order to **aggregate** information about pros and cons of each hotel and produces a **ranking** of hotels taking into account the users interests and the pros and cons of each hotel

*2) the **Reuse** process combines, in some specific way, the (experiential) content of those retrieved cases (and possibly using some domain-specific knowledge as well) in order to achieve a solution for that problem (that specific task to be achieved)*



Task specification

I-day business trip

4-days family vacation

PROS

CONS

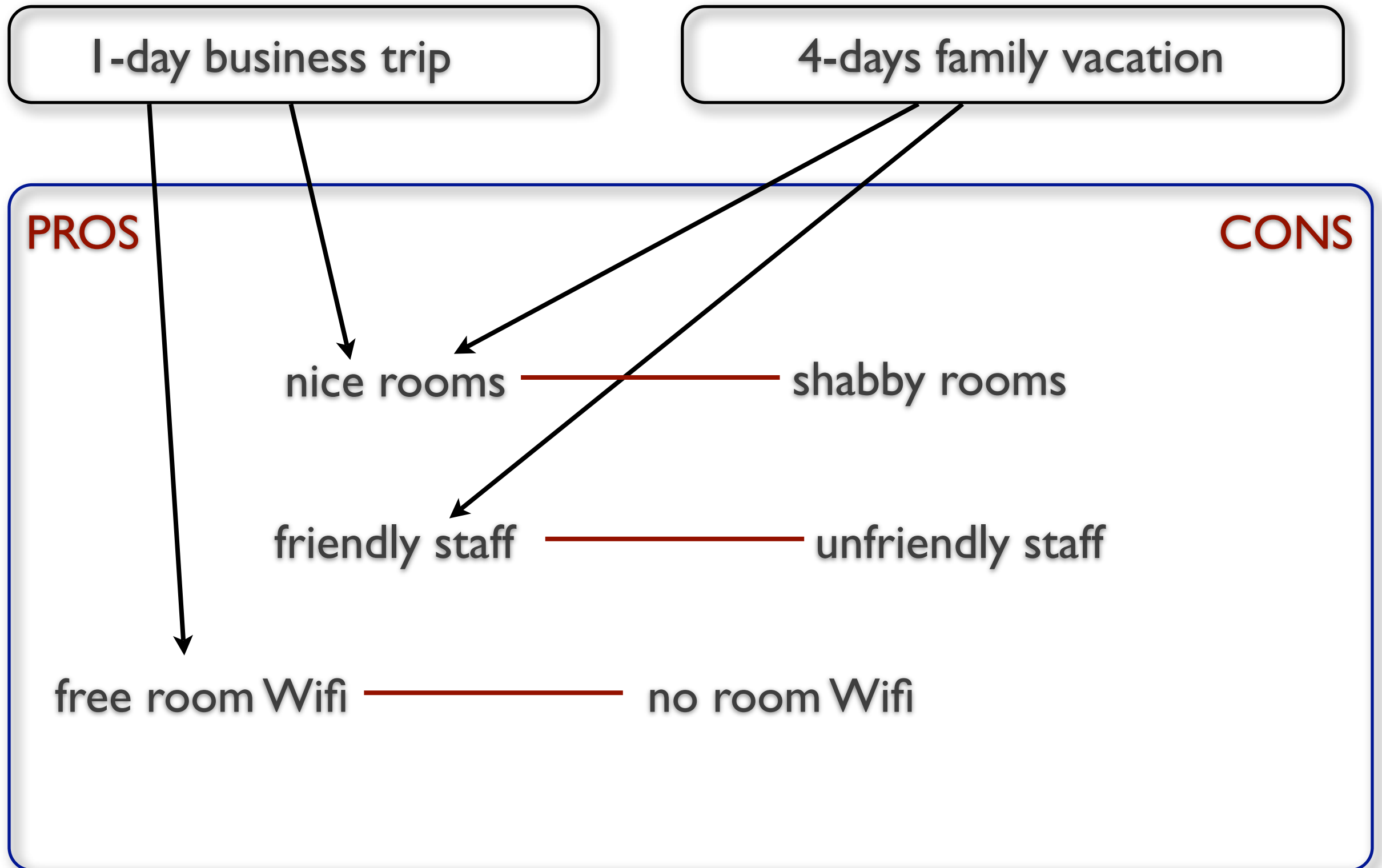
nice rooms — shabby rooms

friendly staff — unfriendly staff

free room Wifi — no room Wifi



Task specification





Ensemble Effect

PROS

CONS

nice rooms — shabby rooms

friendly staff — unfriendly staff

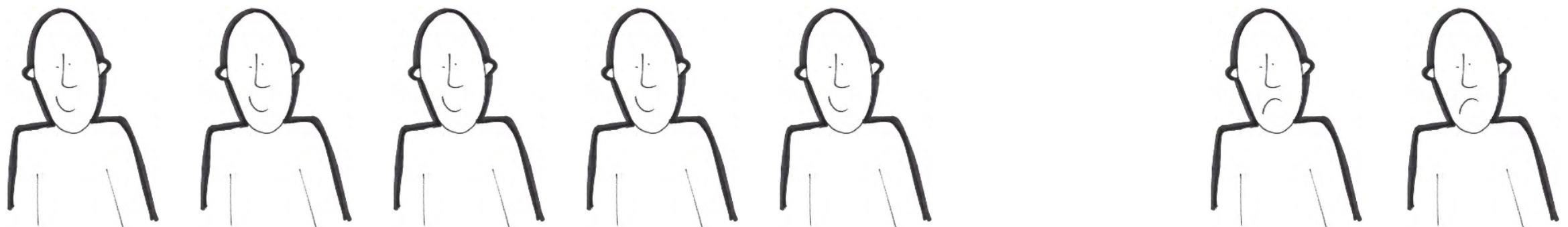


Ensemble Effect

PROS

CONS

nice rooms — shabby rooms
friendly staff — unfriendly staff





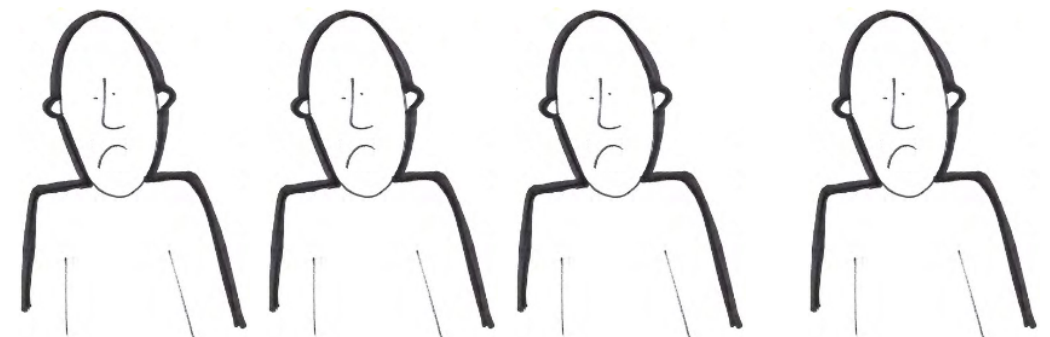
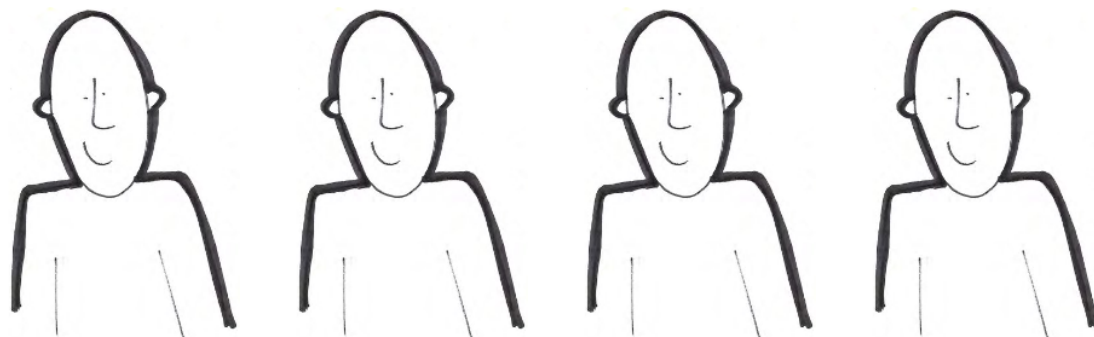
Ensemble Effect

PROS

CONS

nice rooms — shabby rooms

friendly staff — unfriendly staff





Aggregation & Experiences

Ensemble Effect

Machine Learning

Condorcet Jury Theorem

Social Choice

“Wisdom of the Crowds”

J. Surowiecki

Aggregation diminishes error iff individuals are minimally competent and their errors are uncorrelated (they are independent)





Aggregation & Experiences

Ensemble Effect

Machine Learning

Condorcet Jury Theorem

Social Choice

“Wisdom of the Crowds”

J. Surowiecki

Aggregation diminishes error iff individuals are minimally competent and their errors are uncorrelated (they are independent)

Retrieve & Reuse applied to other people's experiences deals with large number of experiences (instead of selecting a few cases); multiplicity of sources indicates the need for the aggregation of experience content

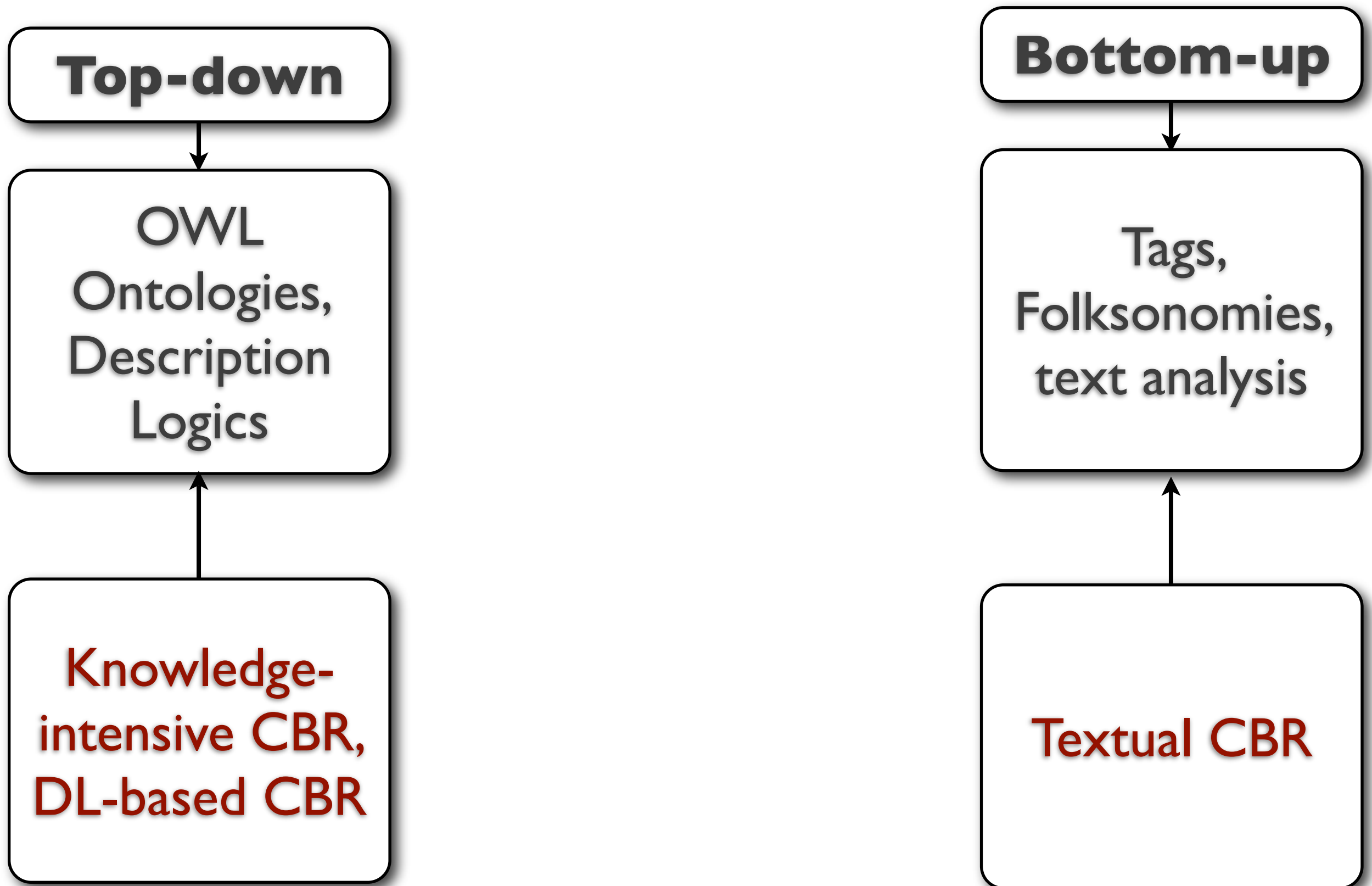




Semantics & Experience

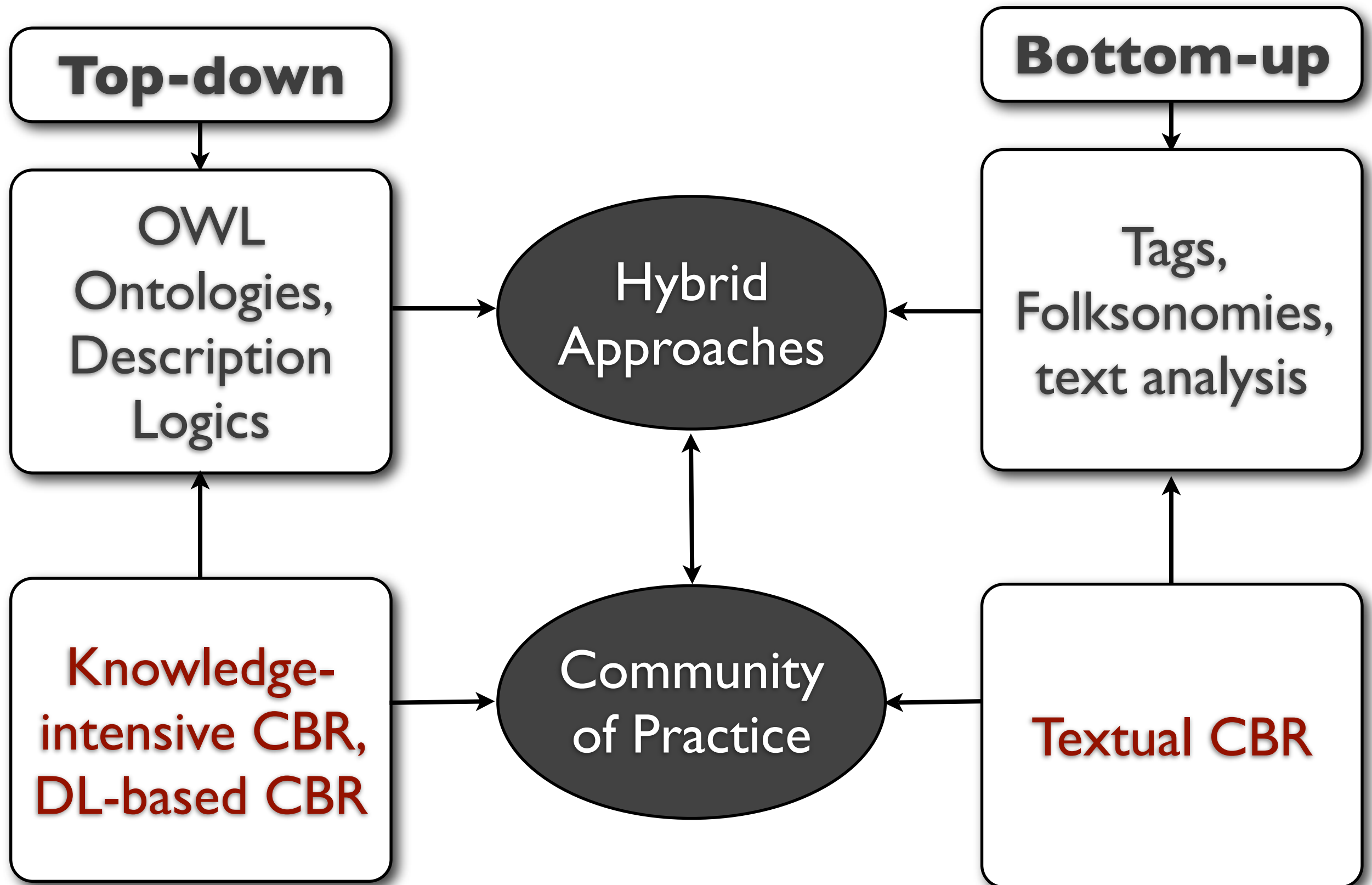


Semantics in CBR



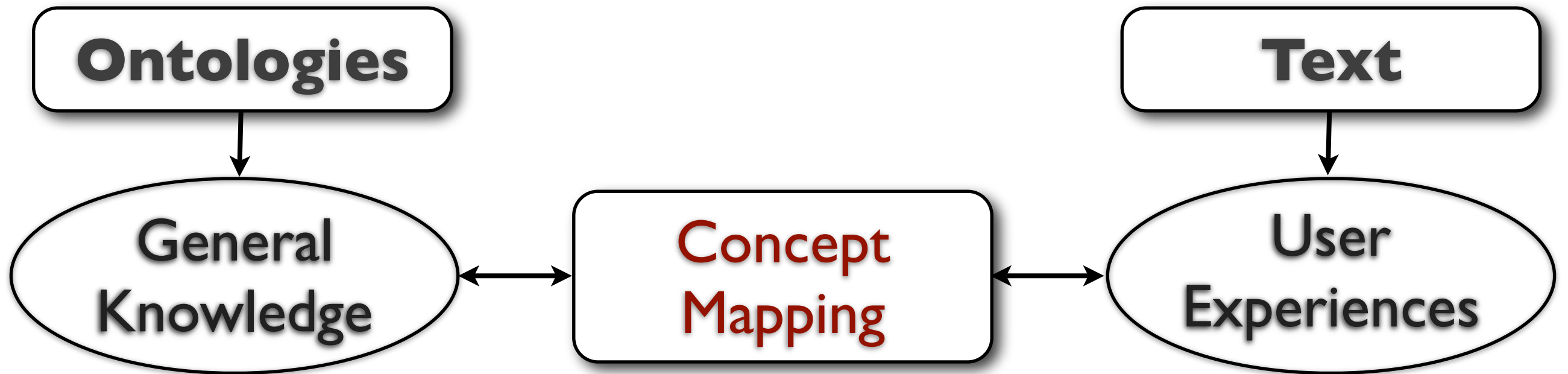


Semantics in CBR





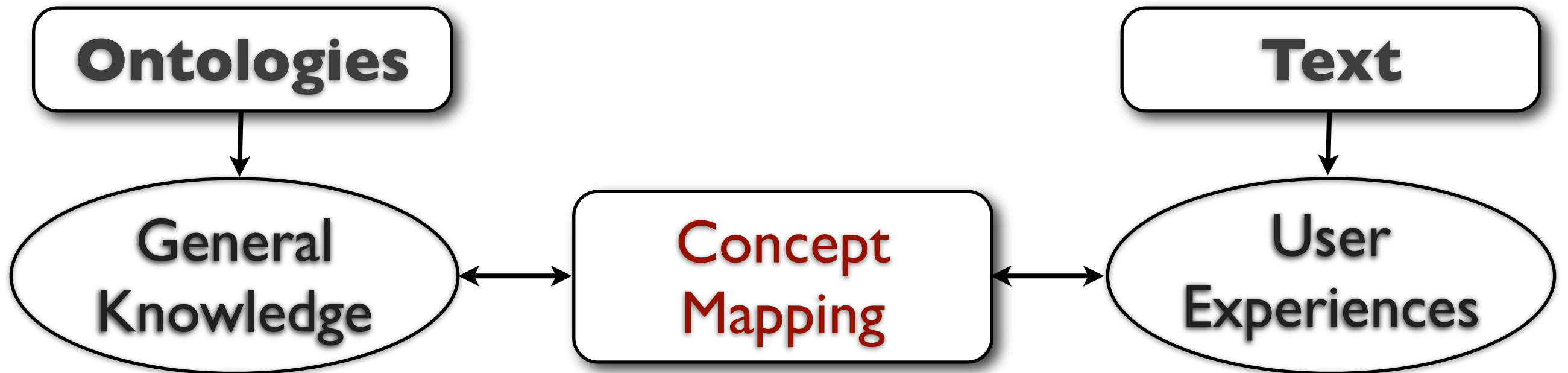
Empirical issues



Tradeoffs are empirical, depends on kinds of application task



Empirical issues



Tradeoffs are empirical, depends on kinds of application task





Forms of Experience



No “cases” as such

No Case as

situation₃

outcome₃

Records of individual experiences

E.g. Hotel client report:

“This is a good hotel because my stay very agreeable”

This is not really case we may find:

user interests
preferences
constraints

selected
hotel



No “cases” as such

No Case as

situation₃

outcome₃

Records of individual experiences

E.g. Hotel client report:

“This is a good hotel because my stay very agreeable”

This is not really case we may find:

user interests
preferences
constraints

~~selected
hotel~~

An **account** of an
experience in a
hotel, with **pros**
and **cons**



No “cases” as such

No Case as

situation₃

outcome₃

Records of individual experiences

E.g. Hotel client report:

“This is a good hotel because my stay very agreeable”

This is not really case we may find:

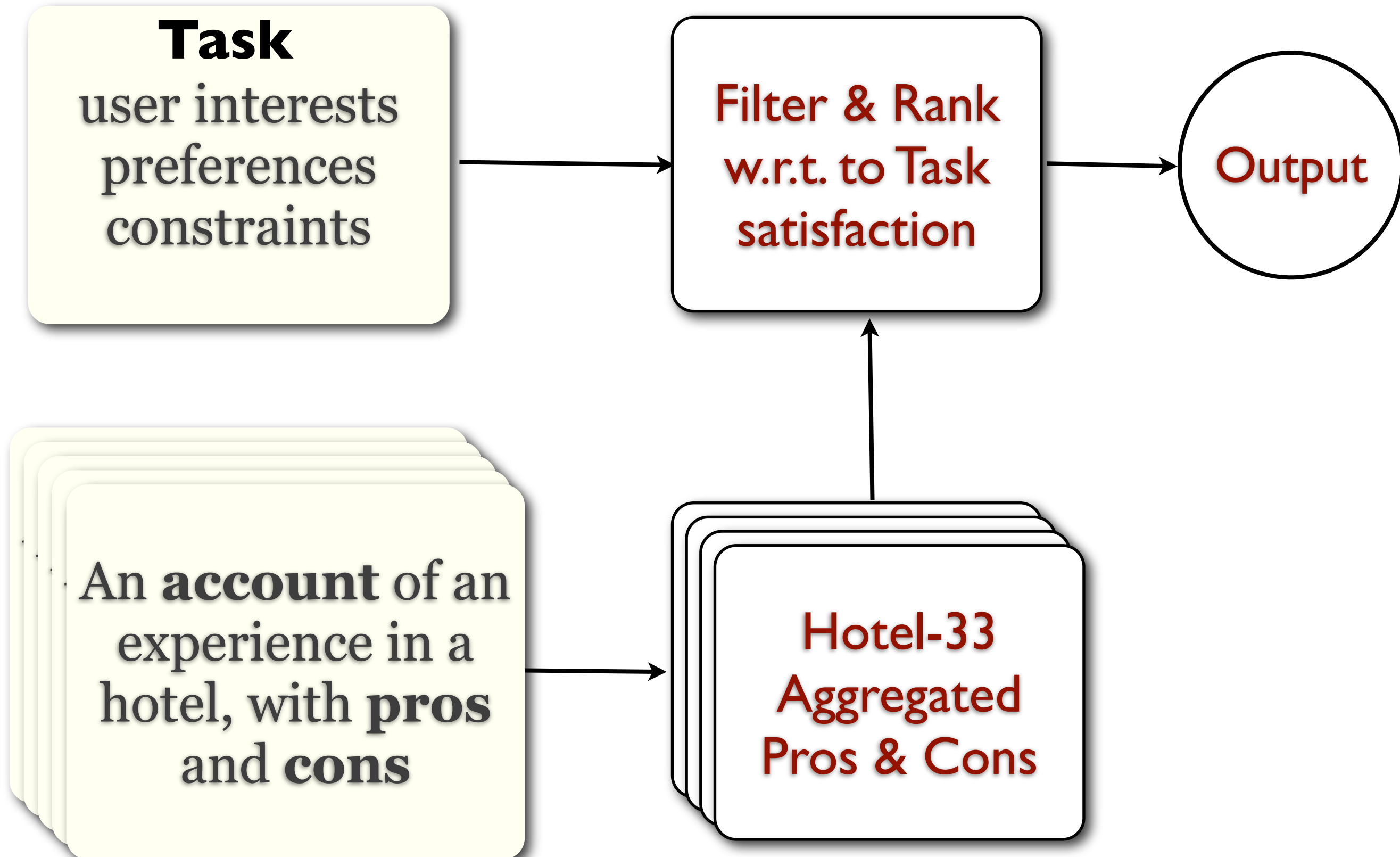
user interests
preferences
constraints

~~selected
hotel~~

An **account** of an
experience in a
hotel, with **pros**
and **cons**



Tacit knowledge





Which Forms of Experience?

According to the form of the solution



Which Forms of Experience?

According to the form of the solution

I) **Classification:** task that selects one or few elements from an enumerated collection of solution elements. Also hierarchical, ranking.



Which Forms of Experience?

According to the form of the solution

1) **Classification:** task that selects one or few elements from an enumerated collection of solution elements. Also hierarchical, ranking.

2) **Regression** a task predicting the numerical value. Case-based interpolation.



Which Forms of Experience?

According to the form of the solution

1) **Classification:** task that selects one or few elements from an enumerated collection of solution elements. Also hierarchical, ranking.

2) **Regression** a task predicting the numerical value. Case-based interpolation.

3) **Planning:** a task building a solution composed by a sequence (or a PO) of actions). Case-base planning; scheduling.



Which Forms of Experience?

According to the form of the solution

1) **Classification:** task that selects one or few elements from an enumerated collection of solution elements. Also hierarchical, ranking.

2) **Regression** a task predicting the numerical value. Case-based interpolation.

3) **Planning:** a task building a solution composed by a sequence (or a PO) of actions). Case-base planning; scheduling.

4) **Configuration:** a task building a solution composed by a network of interconnected solution elements. case-based configuration and design



Which Forms of Experience?

1) **Classification:** task that selects one or few elements from an enumerated collection of solution elements. Also hierarchical, ranking.

METHOD

Pros & Cons Analysis

Applicable to Classification tasks like:

Hotel selection

Digital Camera

B/W Plugin Photoshop

METHOD

METHOD



‘How-To’s as plans

Small plans, recipes or ‘how-to’s are ubiquitous in thematic websites, but they are organized as Q&A, Forum threads, blog entries, etc.

Typical Scenario: User performs S&B to find how to perform a certain effect on a digital photography

Typical Solution: A plan or ‘how-to’ of the form
“*assuming you have Photoshop, you should download this PluginX from this URL, install it and then set it up in beginner mode and you’ll have a good quality B/W image*”



Semi-structured experience

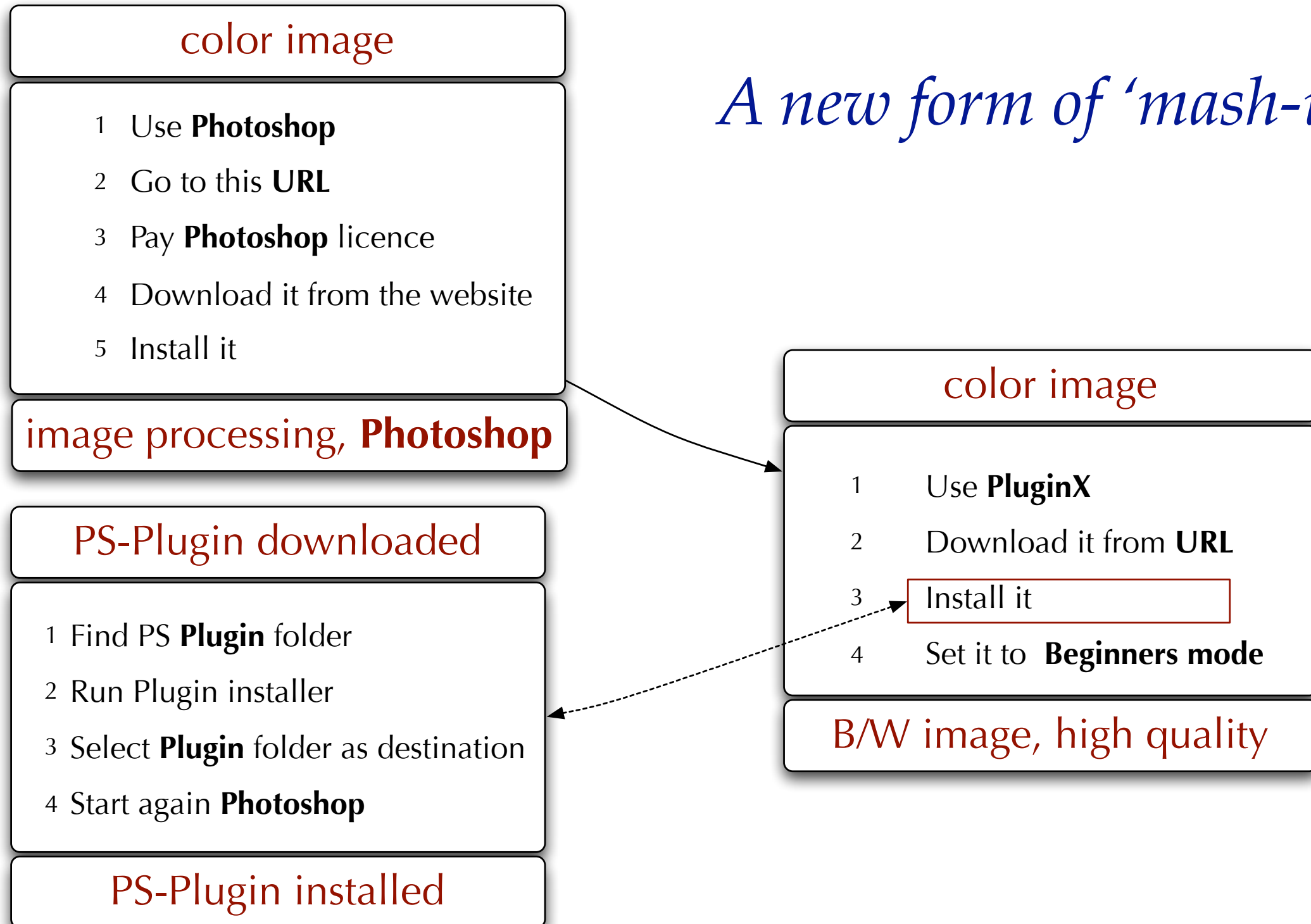
<i>Assumptions</i>	
1	Step description
2	Step description
3	Step description
N	Step description
<i>Effect</i>	

Photoshop, color image	
1	Use PluginX
2	Download it from URL
3	Install it
4	Set it to Beginners mode
B/W image, high quality	



Adapting experience

A new form of 'mash-up'

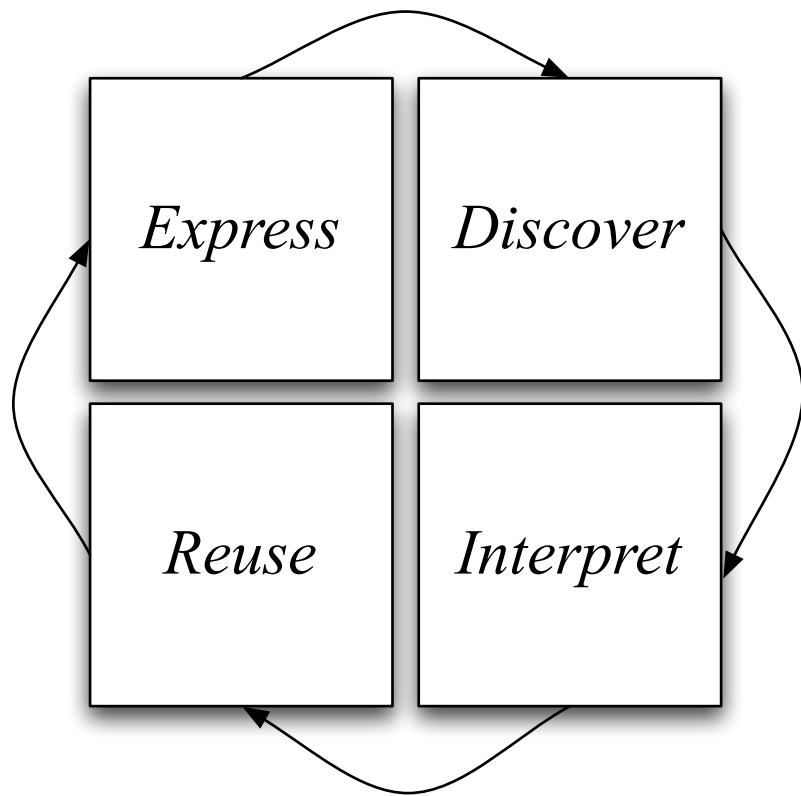




The EDIR Cycle



EDIR cycle



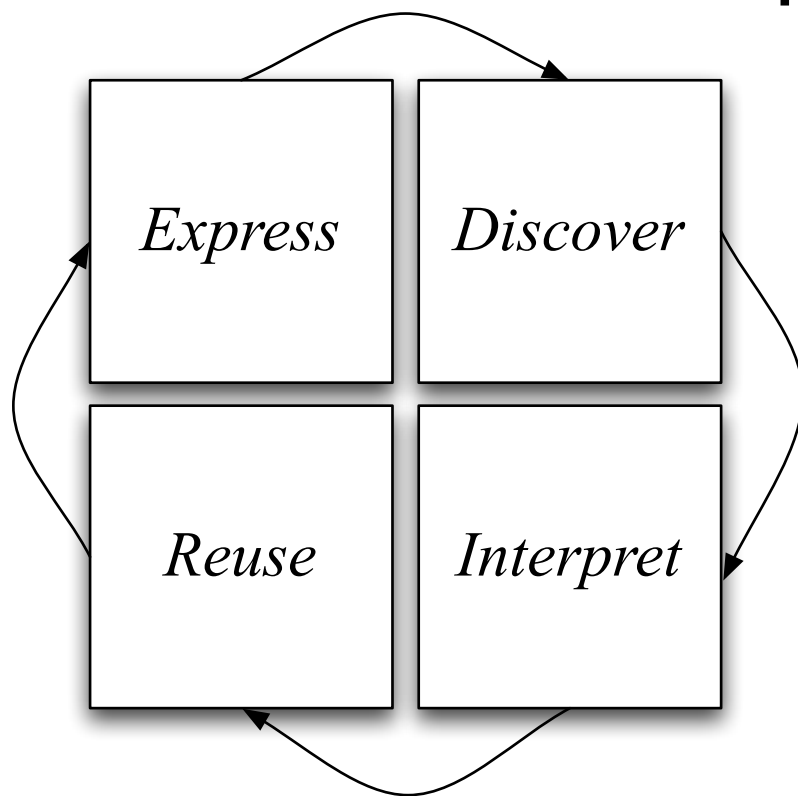


EDIR cycle

Express This process addresses the different ways in which experience can be expressed by a user inside a community of practice.

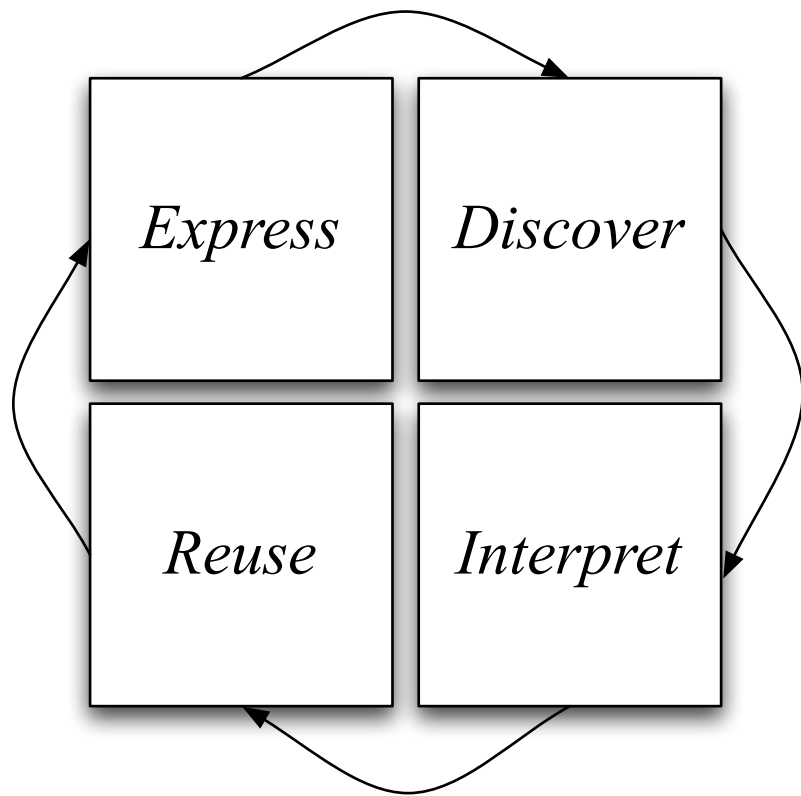
Free, semi-structured and ontology-based templates for specific forms of experience and application domains need be developed and tested.

Research goal: finding a trade-off that:
(a) allows sufficient structuring of the expressed experiences for automated analysis and
(b) feels as a natural and unobtrusive way to express experiences for the users in a community of practice.





EDIR cycle



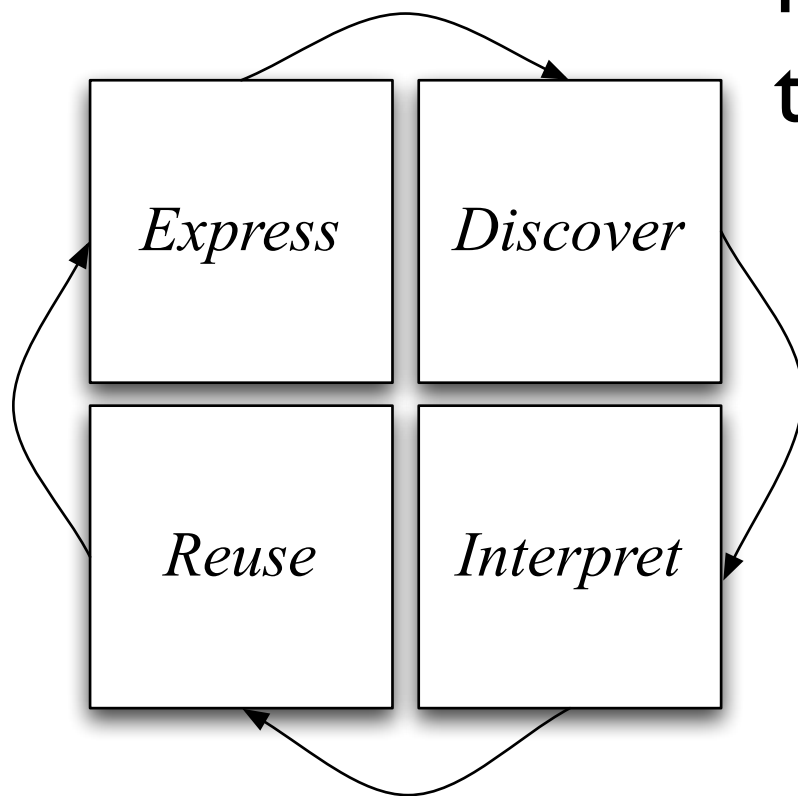


EDIR cycle

Discover This process addresses the different ways in which specific experiential content is recognized and retrieved as possibly relevant to a given query posed by a system user

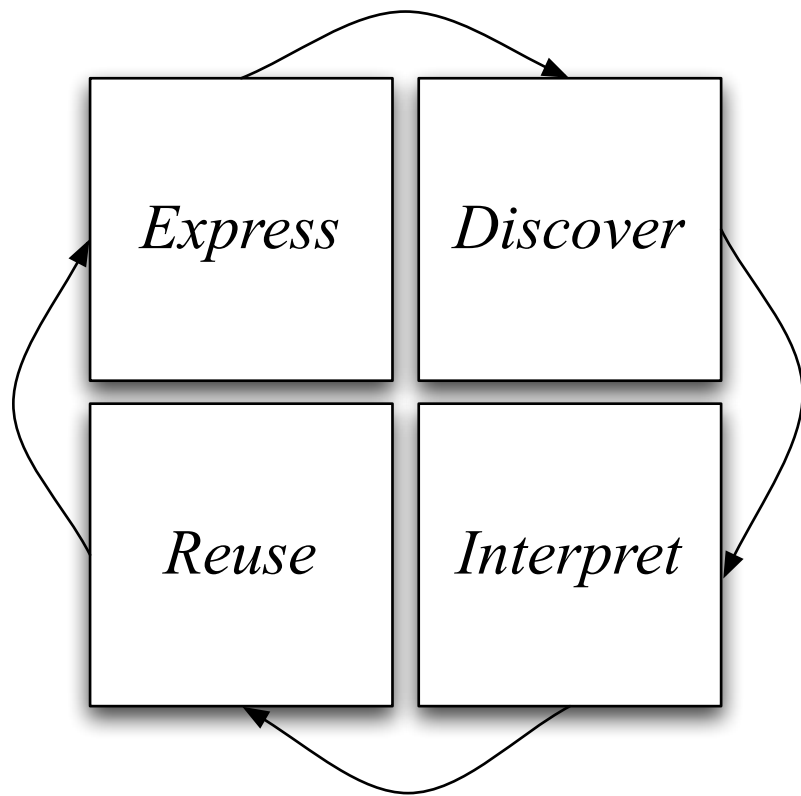
Research goal: how to extend CBR retrieval techniques to work on experiential content integrating semantic web and/or bottom-up semantic analysis.

The conditions under which the Discovery process has to work requires a fast and possibly shallow analysis of large quantities experiential reports; the expected output is a moderately-sized collection of experiences that are (likely) relevant to the current query





EDIR cycle

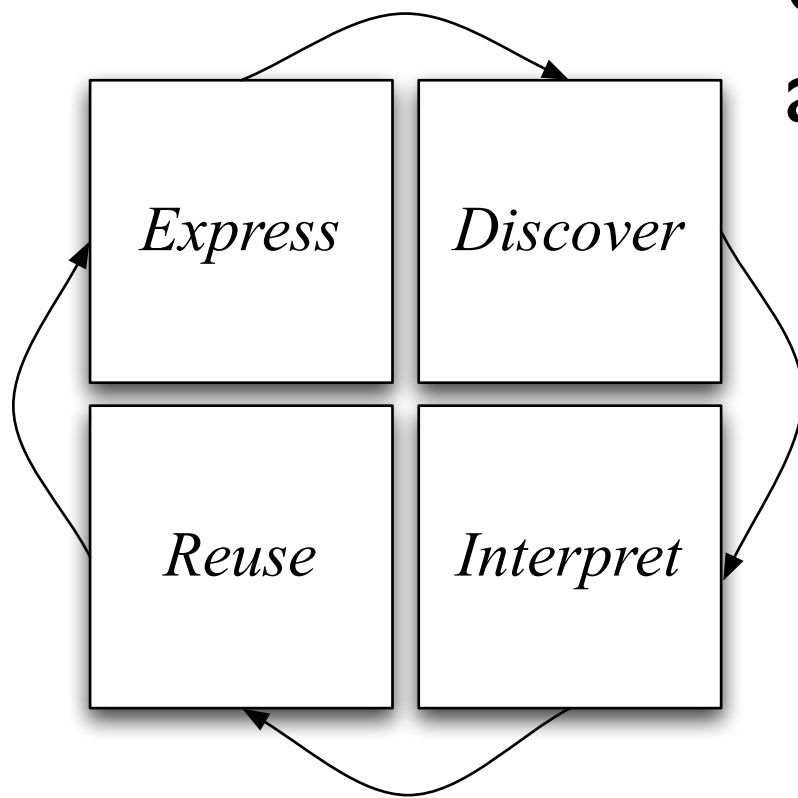




EDIR cycle

Interpret This process addresses the different ways to build semantic interpretations of the discovered experiences. Semantics are assumed local to a community of practice

This interpretation can be understood as a more in-depth analysis of the experiences selected by the Discovery process using the semantic model of the community of practice and the available domain knowledge.

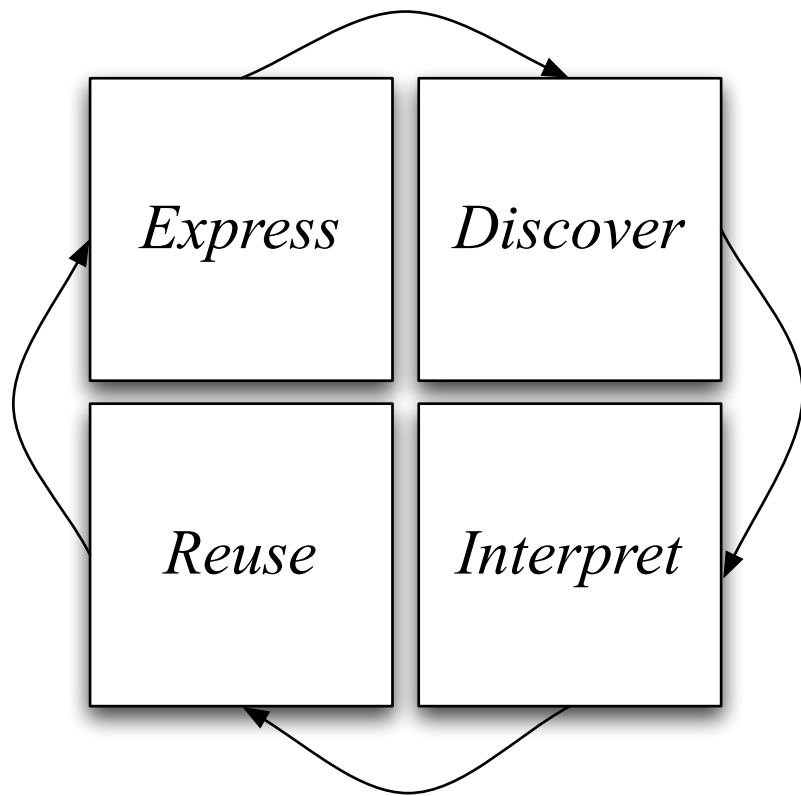


Several transformations are envisioned in the Interpret process:

- (a) eliminating a subset of discovered experiences as non-relevant;
- (b) transforming discovered experiences into a new canonical representation;
- (c) translating discovered experiences into a canonical vocabulary coherent with the one used to build the final users queries



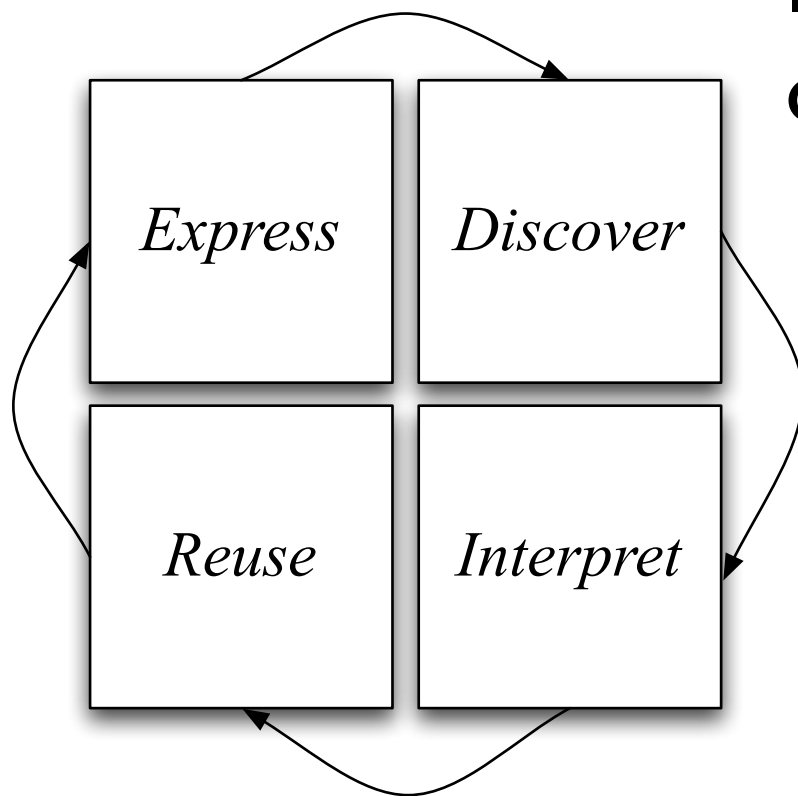
EDIR cycle





EDIR cycle

Reuse This process addresses the different ways in which the experiential content provided by the Interpret process is used to achieve the goals of a user (task specification)



Methods for reuse can vary:

- (a) CBR adaptation techniques
- (b) Aggregation operations exploiting the “ensemble effect”

Modalities in reuse can vary:

- (a) automated solution adaptation for user query
- (b) semi-automated reuse
- (c) reuse left to the user



EDIR vs S&B

Task specification requirement

It's the goal to be achieved by Reuse: otherwise experience cannot be really “reused” and user support would be very limited

Content Organization and Form

Now form are hyperlinked documents, and organization is not based on content but metaphors like forums, Q&A, diaries, etc. A particular type o content, experiential knowledge, may have a few **forms** that can be **expressed** and **represented** in such a way that more powerful ways of **analyzing, organizing, retrieving, and reusing** can be developed.



Task spec

No “problem”
as such

situation₃

outcome₃



- 1) user selects features in a dialog
- 2) “business trip” typical features

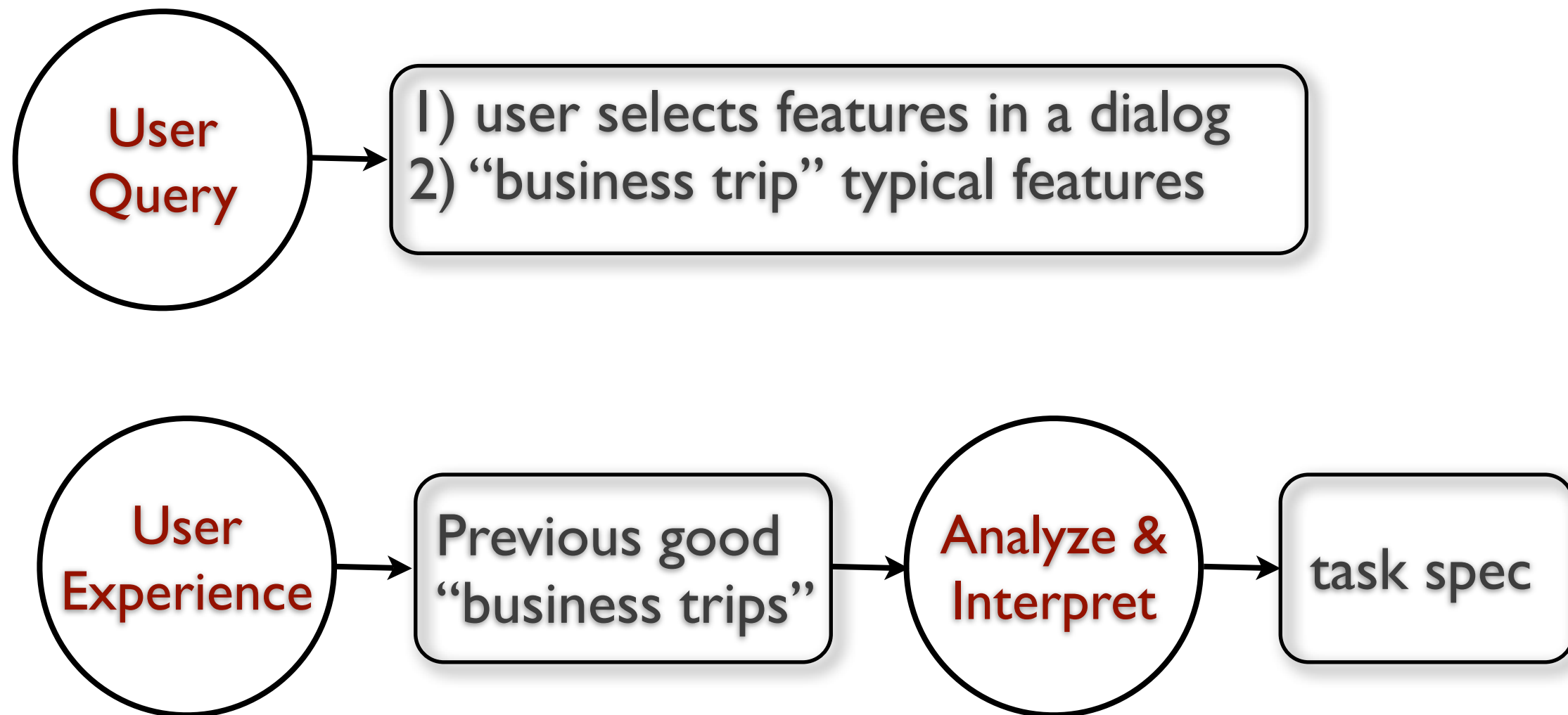


Task spec

No “problem”
as such

situation₃

outcome₃





Conclusions



Conclusions

The challenges and assumptions I've made boil down to:

*There is such a thing as
'experiences', that can
be studied as such, and
they are a particular
kind of content*

*Experiential knowledge
of others can be
acquired and then
reused for new people's
purposes*

*and people are
persistently expressing
them on the web*

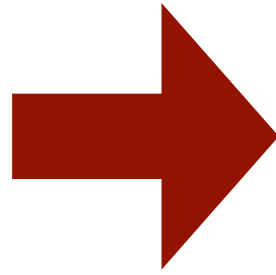
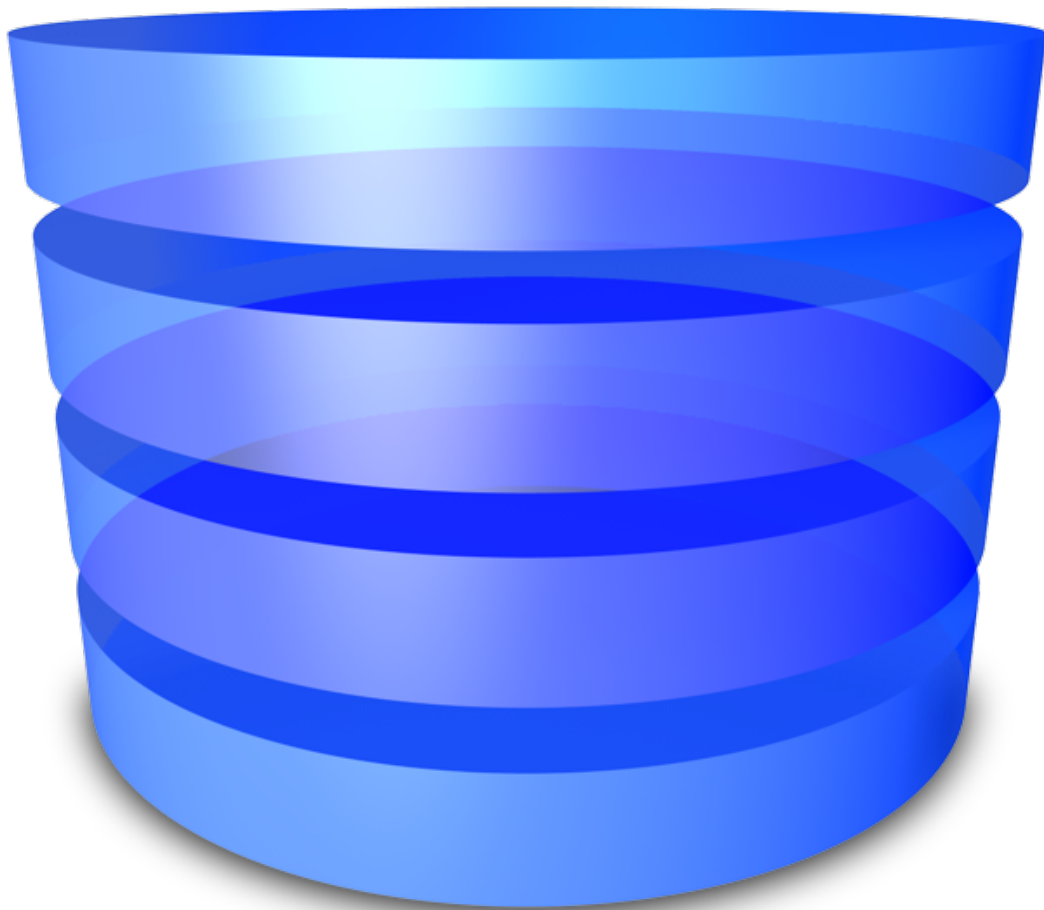
*and people are
persistently trying to do
that on the web*



Conclusions



Conclusions

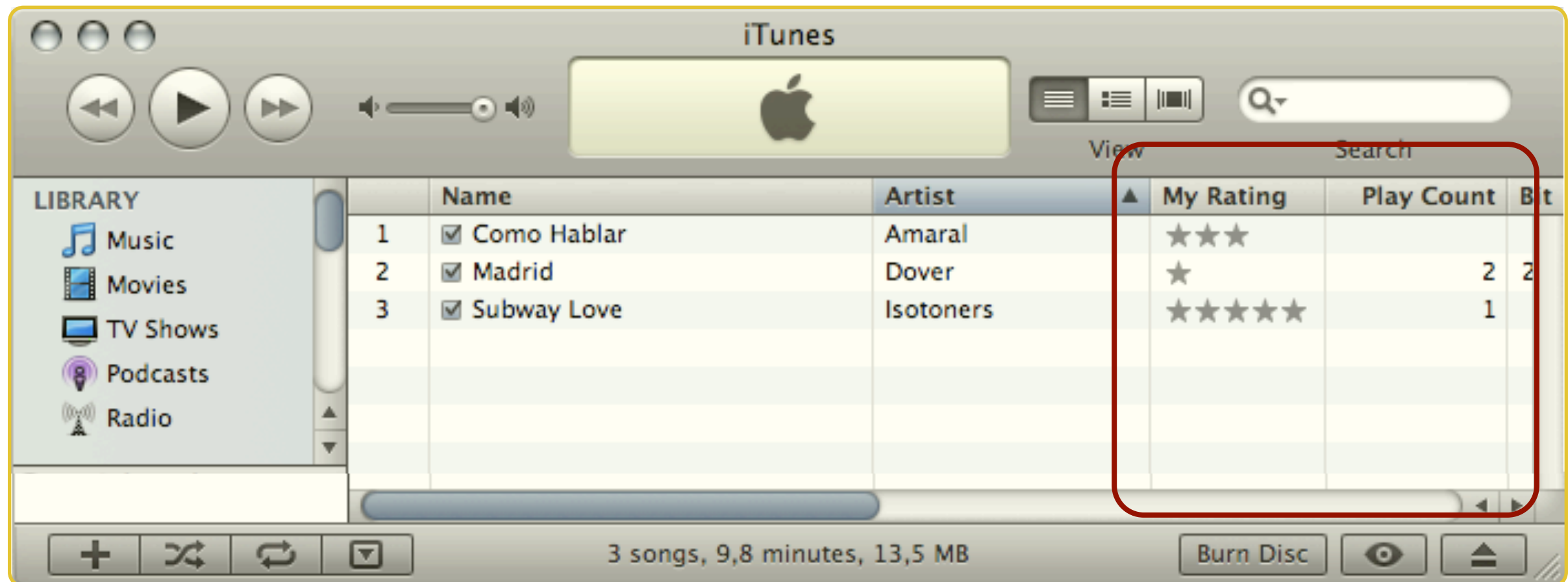




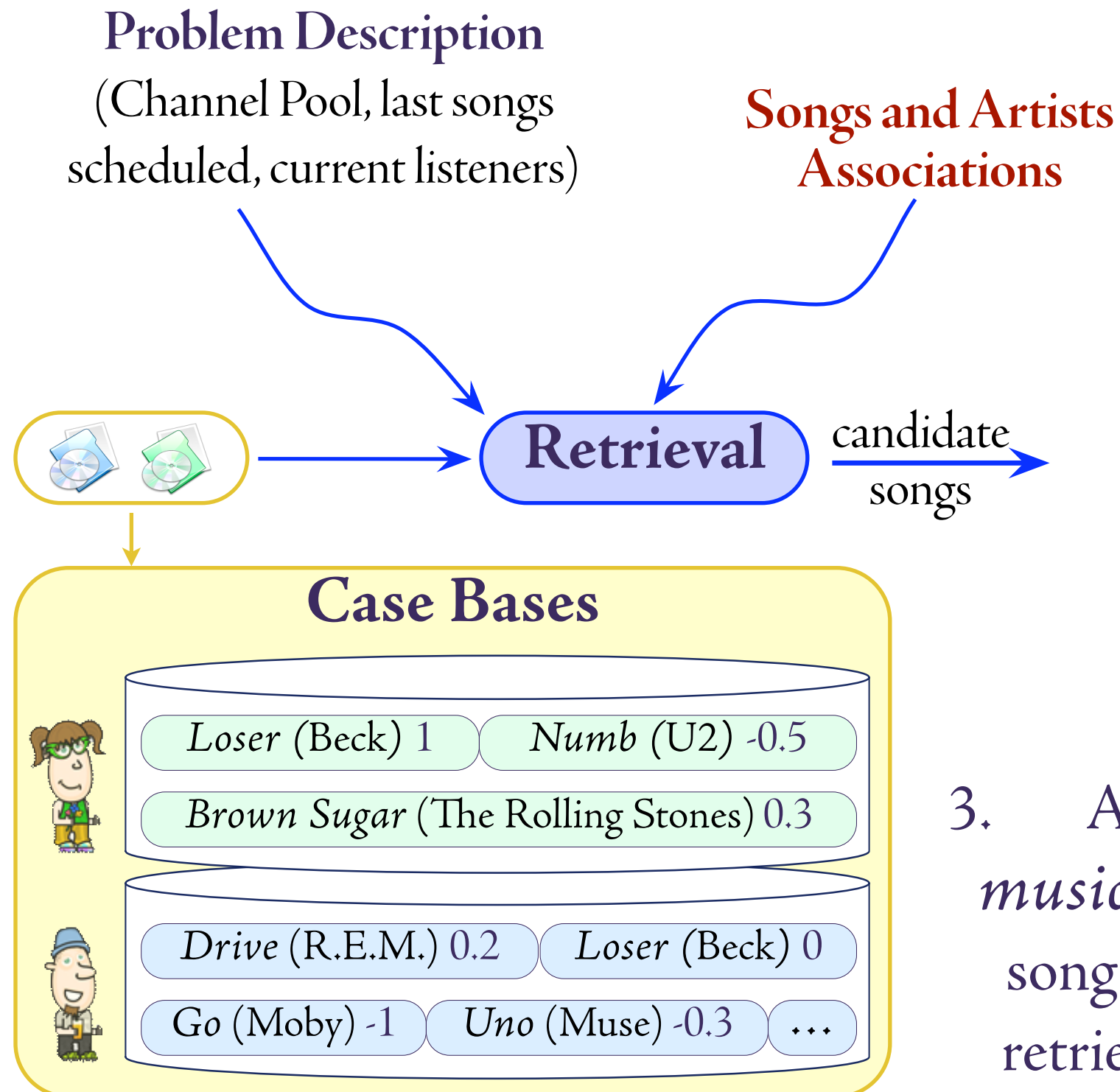
Thanks!

The Participants' Case Bases [2/3]

We assess the **preference degree** of a participant for a song *contained* in her library combining the **rating** assigned to the song and the **number of times** it was listened to:



The Retrieve Process



3. A subset of candidate songs *musically associated* with the last song scheduled on the channel is retrieved from the Channel Pool.

The Reuse Process

