

Ontology-Based Recommender System for Information Support in Knowledge-Intensive Processes

Yordan Terziev, Marian Benner-Wickner,
Tobias Brückmann, Volker Gruhn

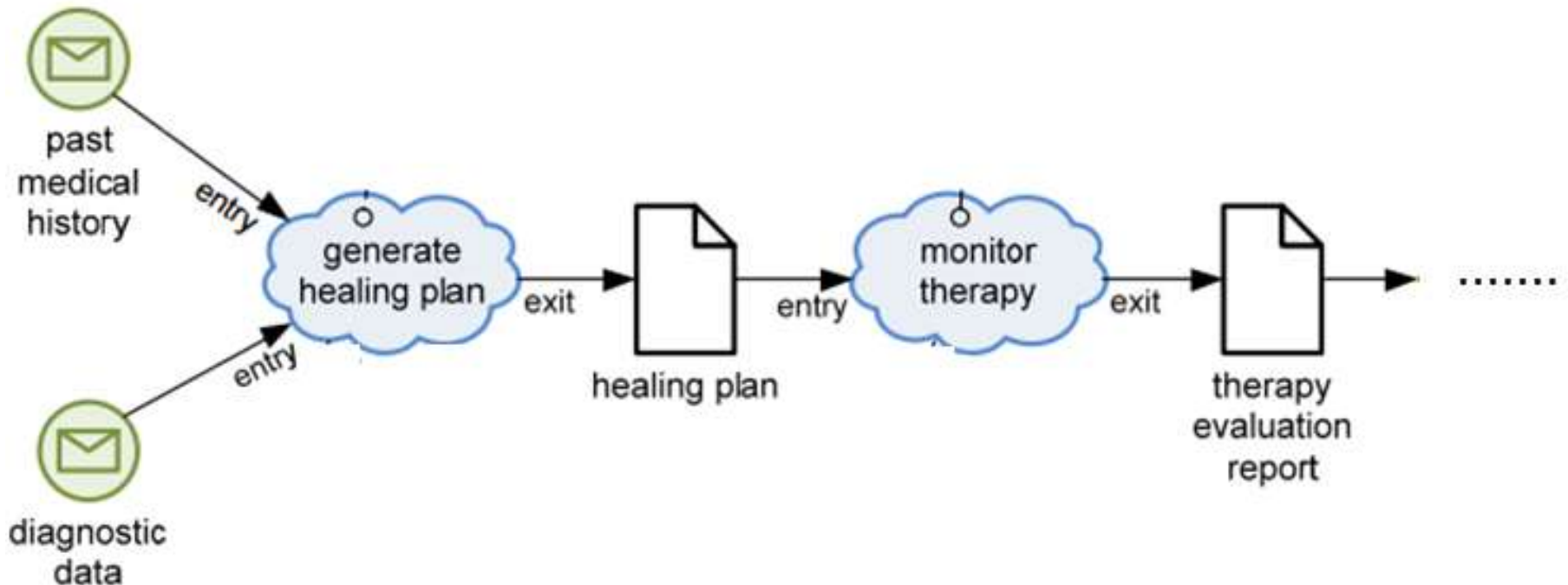
21.10.2015

Overview

1. Why we need better information support for knowledge workers?
2. The problems of supplying case related information.
3. Conceptual foundations
4. Work contributions
5. Evaluation & Results
6. Conclusion

Who are the knowledge workers?

- **Focus:** Case managers executing repeatable tasks
- **Example:** Rehabilitation manager



Source: [4]

Why do we need better **information support** for knowledge worker?

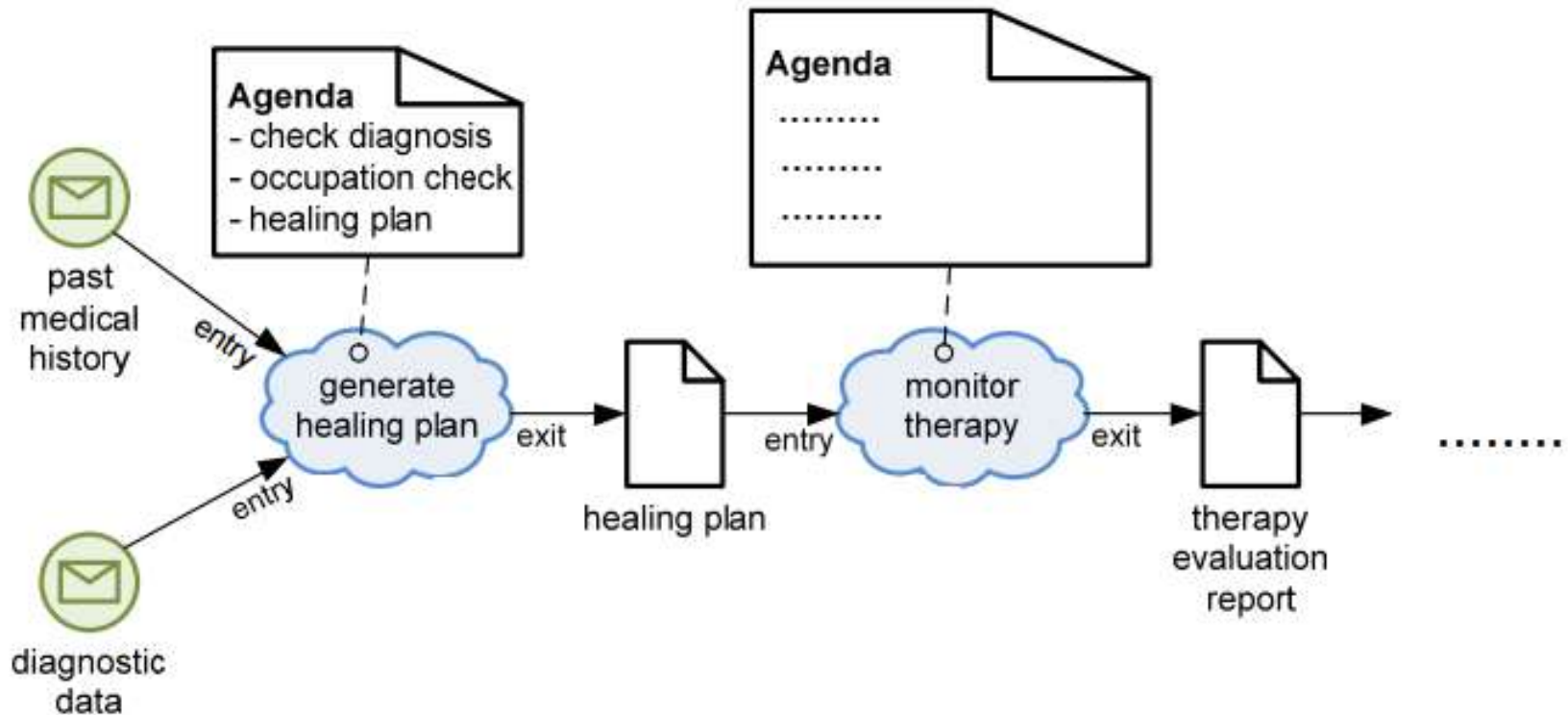
- **Motivation – IDC Study [10]**
 - **9.5 hours per week, searching** for information
 - A third of the time searching is fruitless.
- **Between 25 and 45 % of the workforce are knowledge worker [5]**
- **Problems**
 - Information is stored **independently from the goal for which it is used or created**
 - Knowledge workers are **unaware of relevant information already gathered in similar cases by colleagues.**

Conceptual Foundations

Agenda-driven Case Management (adCM)

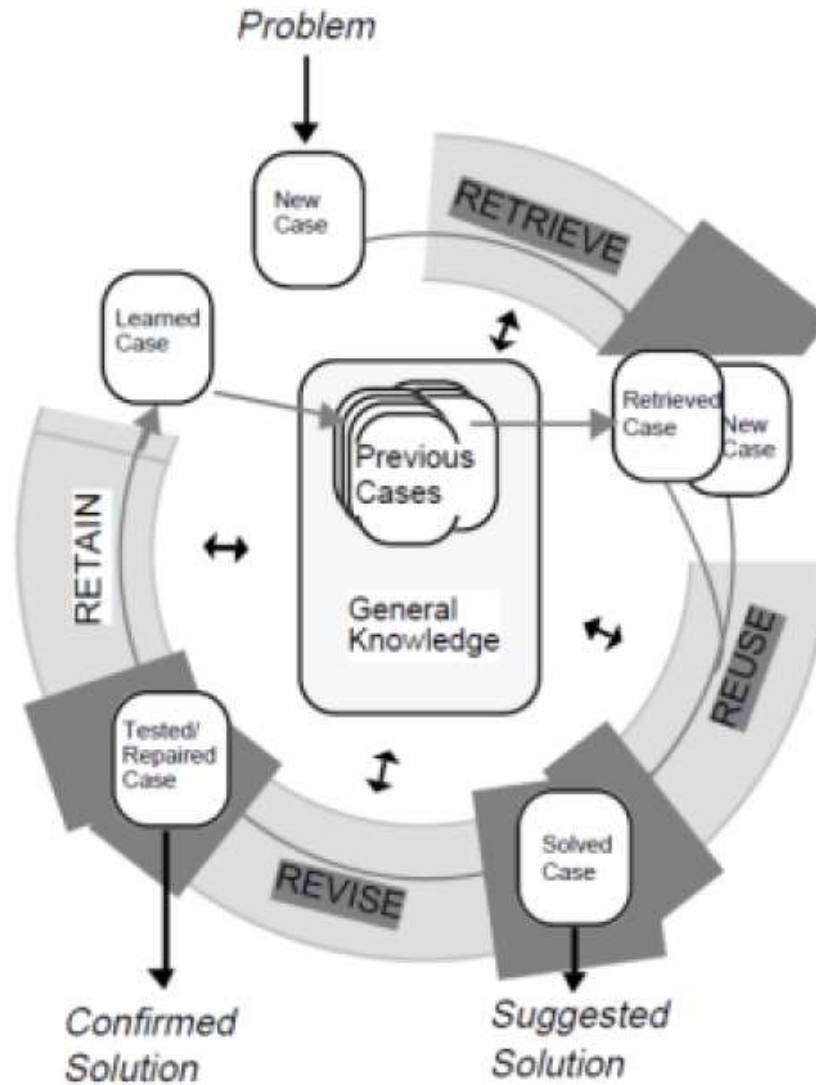
Goal: Support unstructured process parts occurring within structured processes

Provides: Hierarchical structuring of information and tasks within a case, annotations, templates, etc.



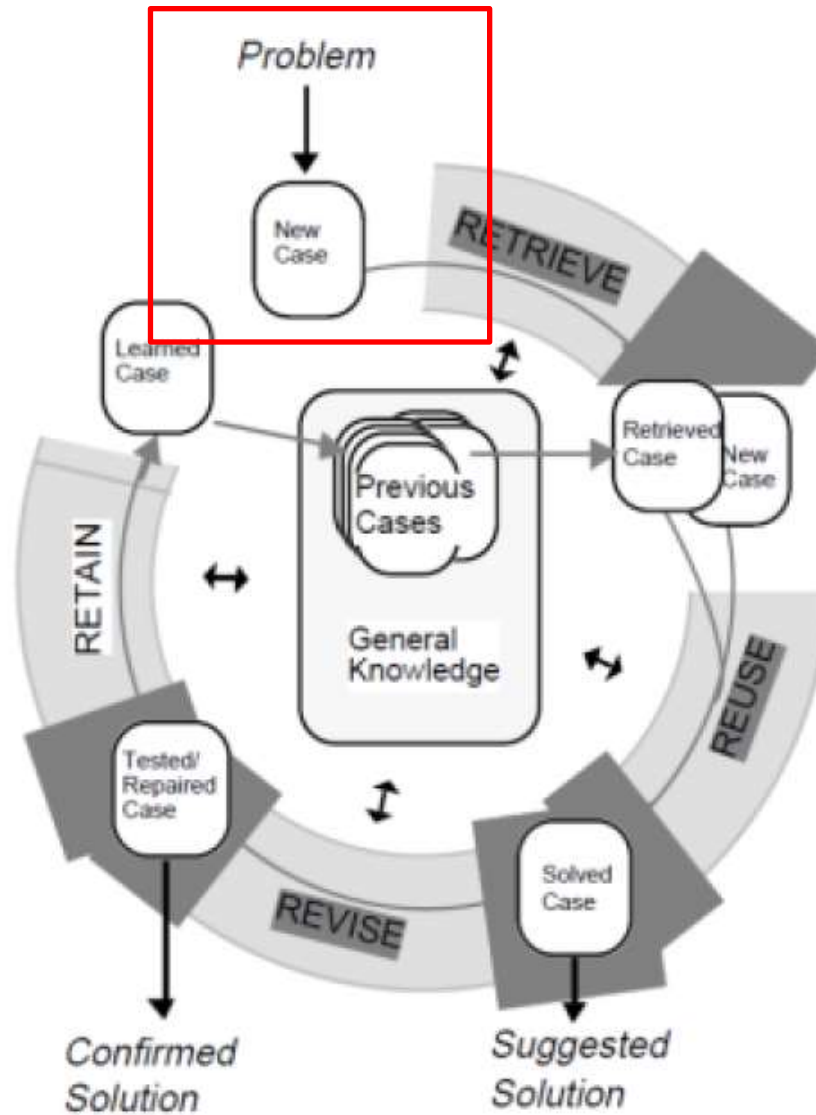
Source: [4]

Case-Based Reasoning (CBR)



Source:[1]

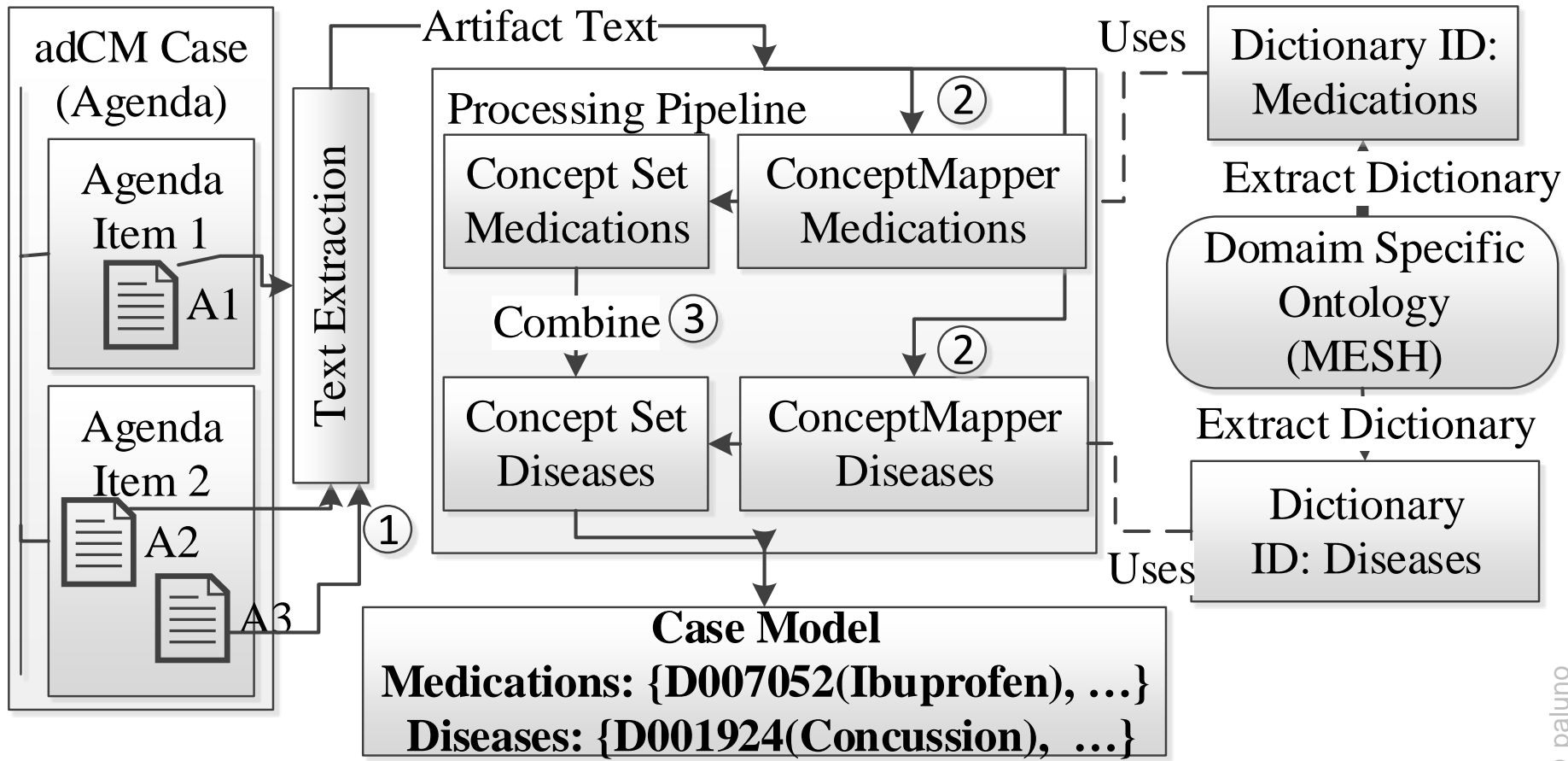
Ontology-Based Recommender System

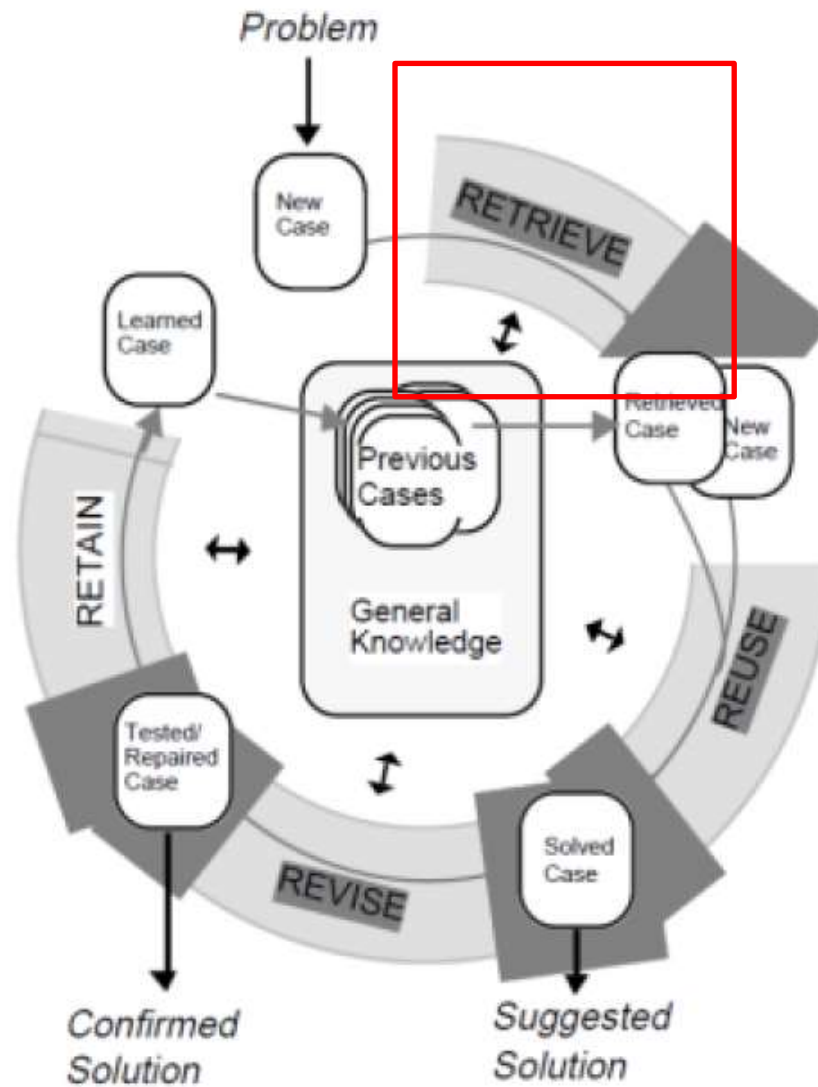


Source:[1]

Ontology Based Case Model

Goal: restrict model to it's most important features





Source:[1]

Retrieve: Case Similarity Measures

Goal: Find similar adCM cases in the case base

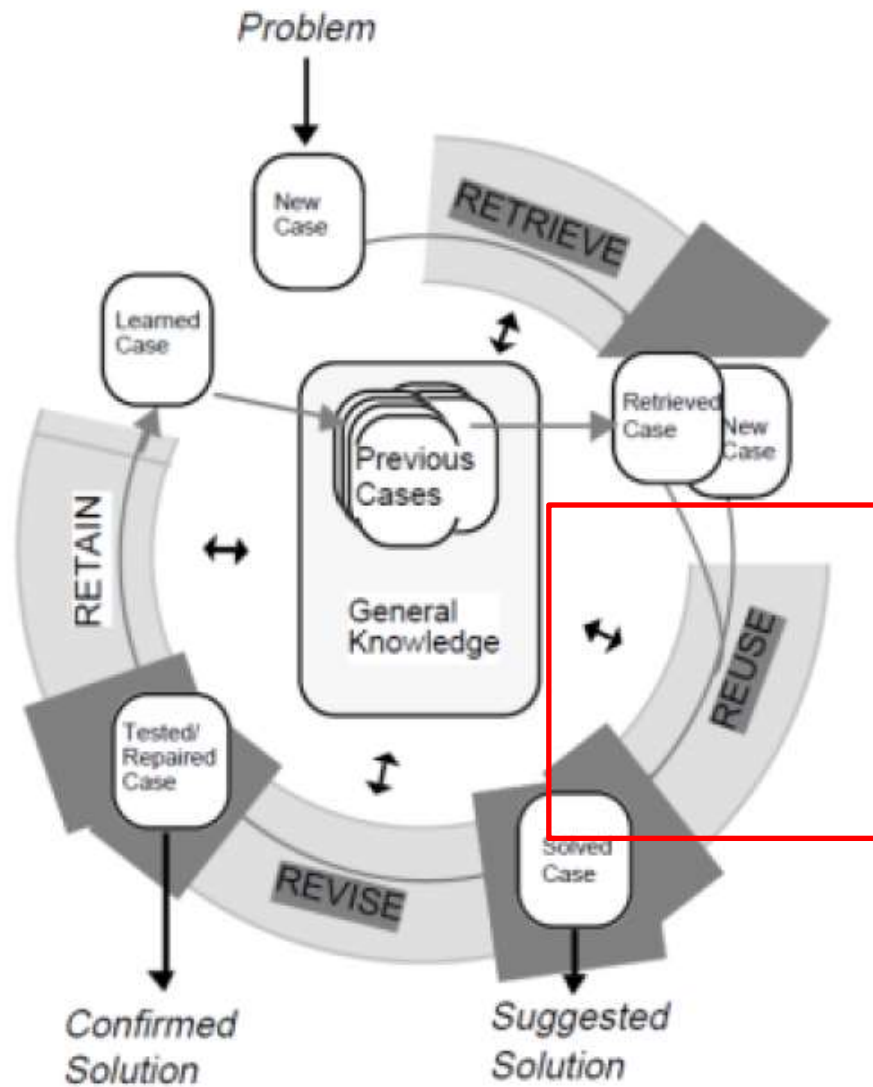
- Use the concept vector and Vector Space Model (VSM) [15]

- Cosine Similarity: $Sim_{cos}(d_1, d_2) = \frac{\vec{V}(d_1) \cdot \vec{V}(d_2)}{|\vec{V}(d_1)| \times |\vec{V}(d_2)|}$

- Concept Frequency and Inversed Case Frequency, **based on:**

- $TF - IDF(t, d, D) = TF_{t,d} \times IDF(t, D)$

- $IDF(t, D) = \log \frac{|D|}{|\{d \in D : t \in d\}|}$

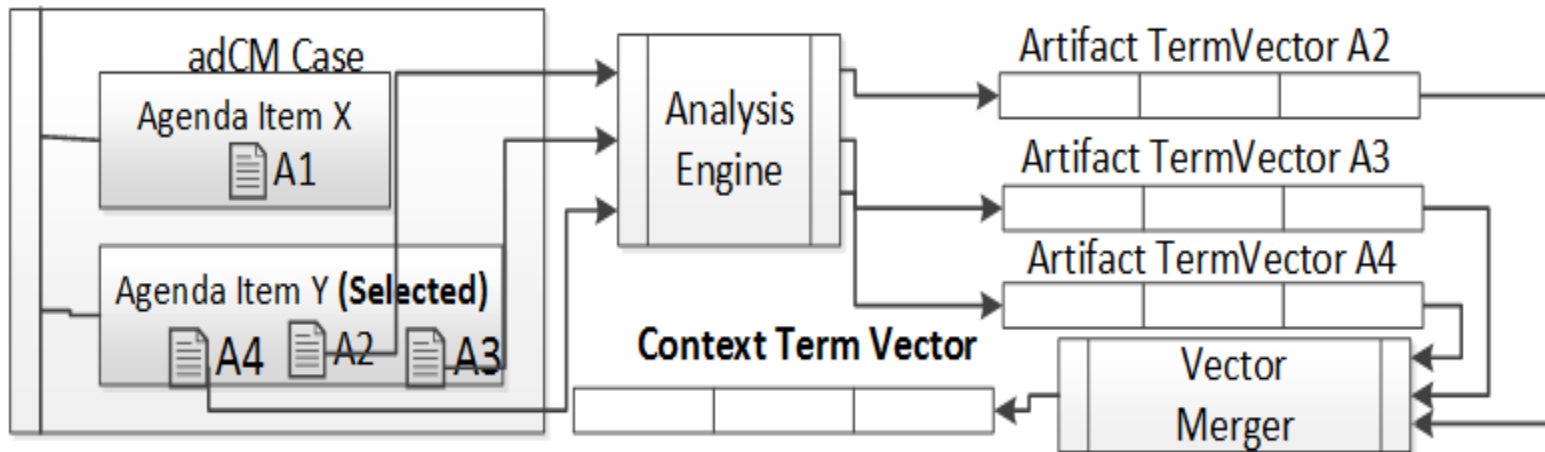


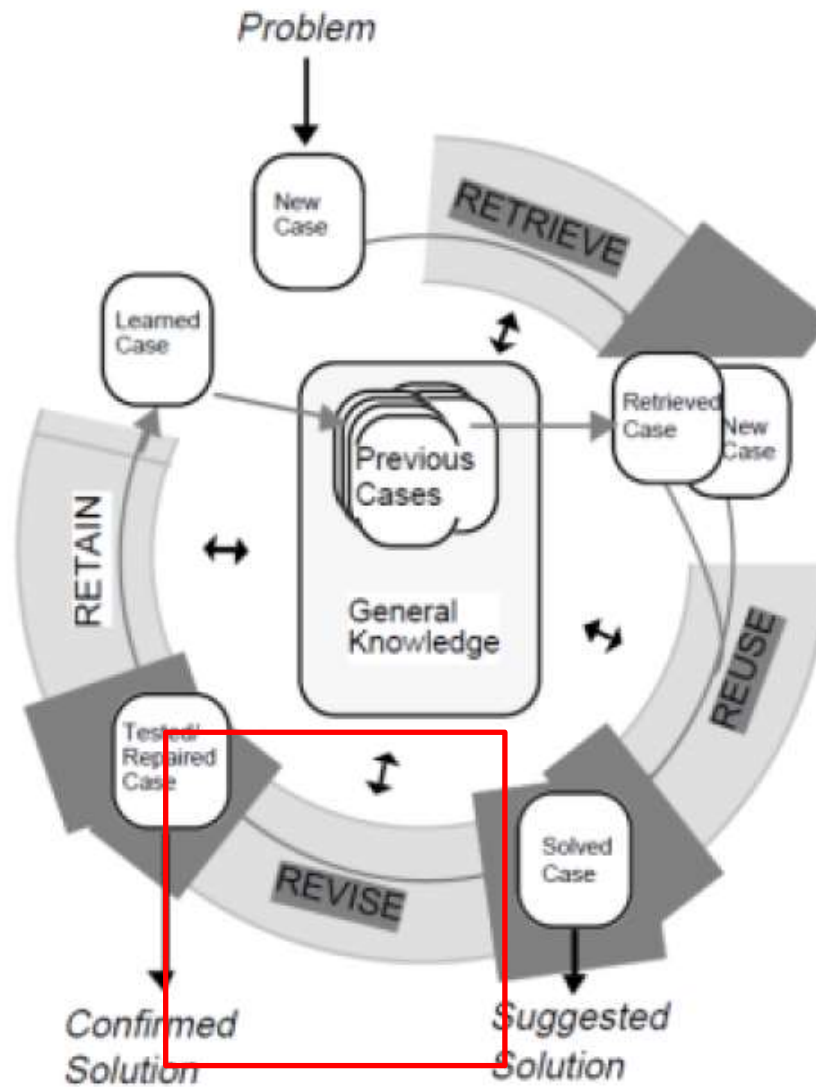
Source:[1]

Reuse: Case Adaptation

Goal: Adjust the retrieved cases to the current case and task of the knowledge worker

1. Remove all artifacts already contained in the current case
2. Prioritize artifacts by currently selected agenda item
3. Prioritize previously **recommended and used** artifacts



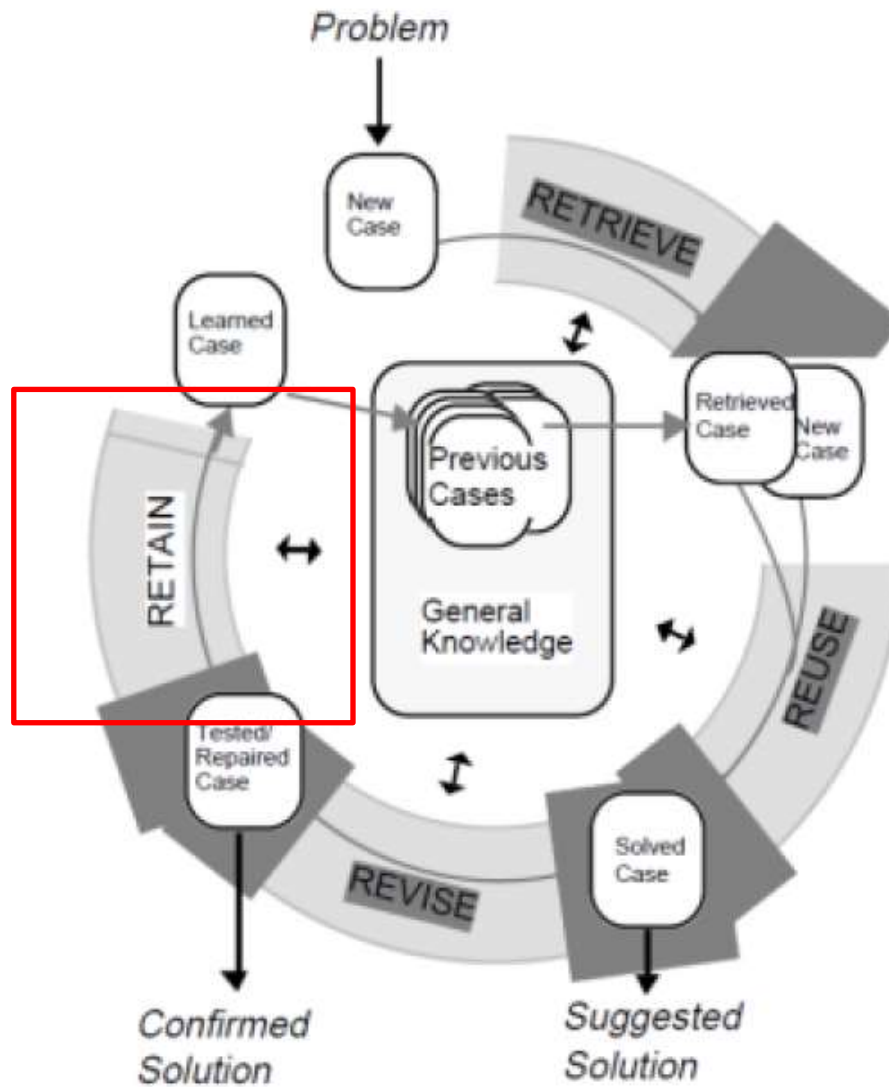


Source:[1]

Revision

Goal: Learn from the user interaction with the recommendations

- **Artifact added to agenda item**
- **Artifact opened**
- **Artifact not selected**

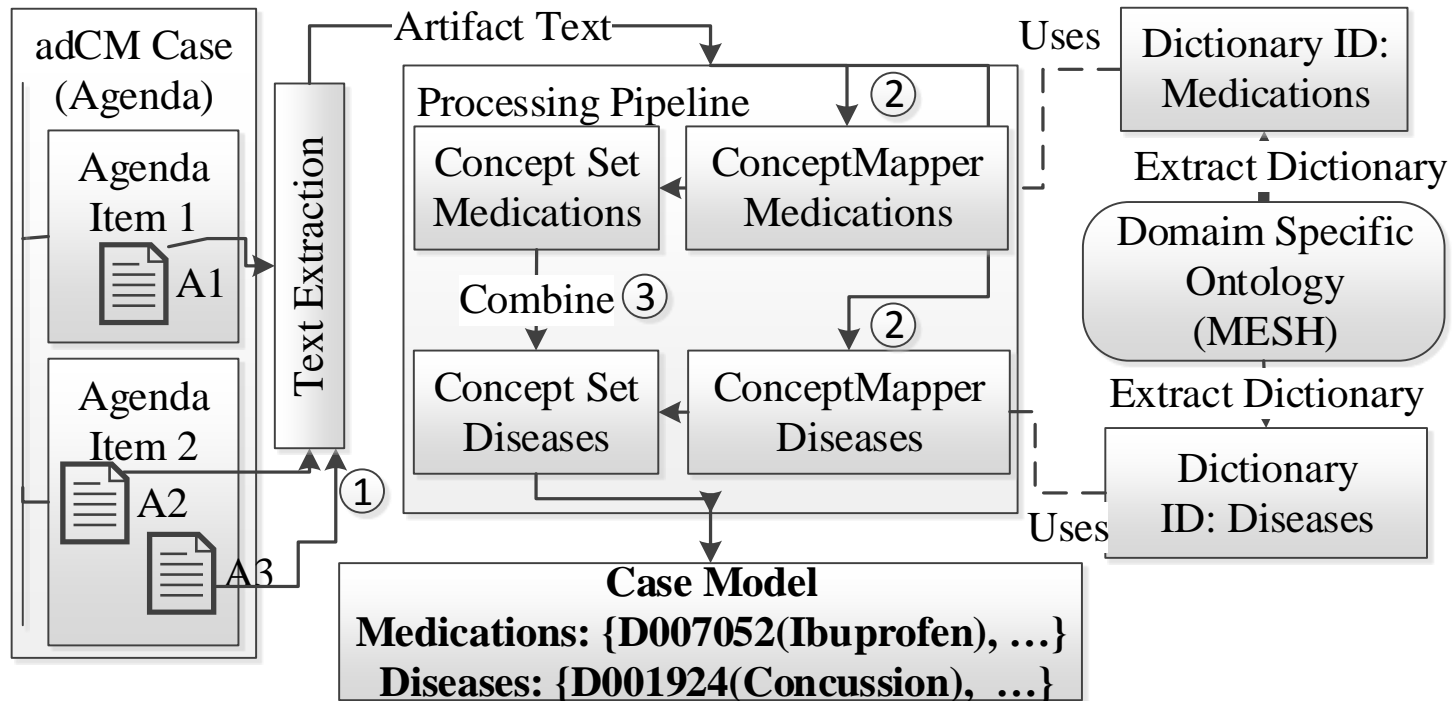


Source:[1]

Retain

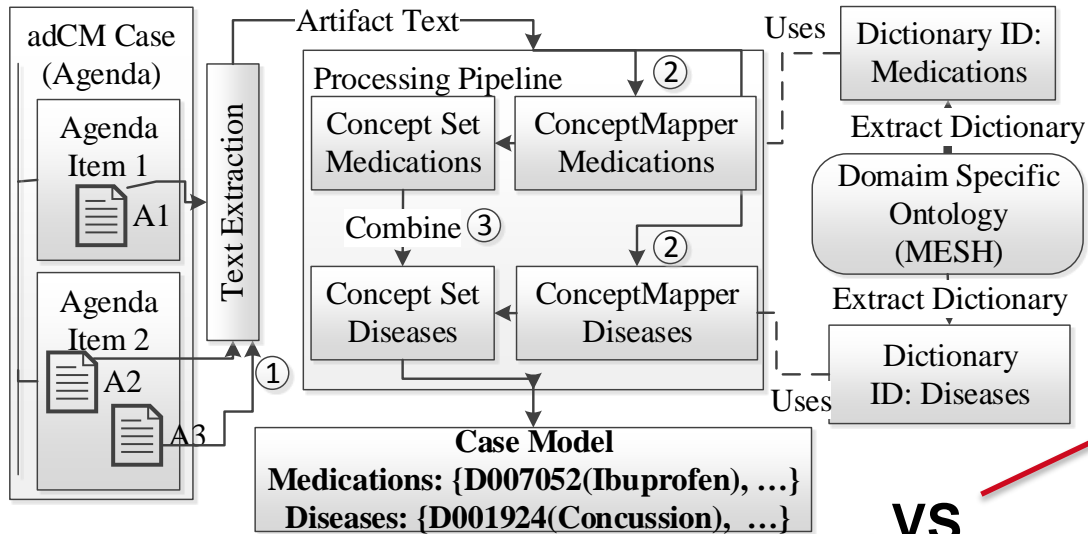
Goal: Process and store completed cases in the case base

- Store statistical distribution of concepts
- Store timestamp (newer cases get prioritized)

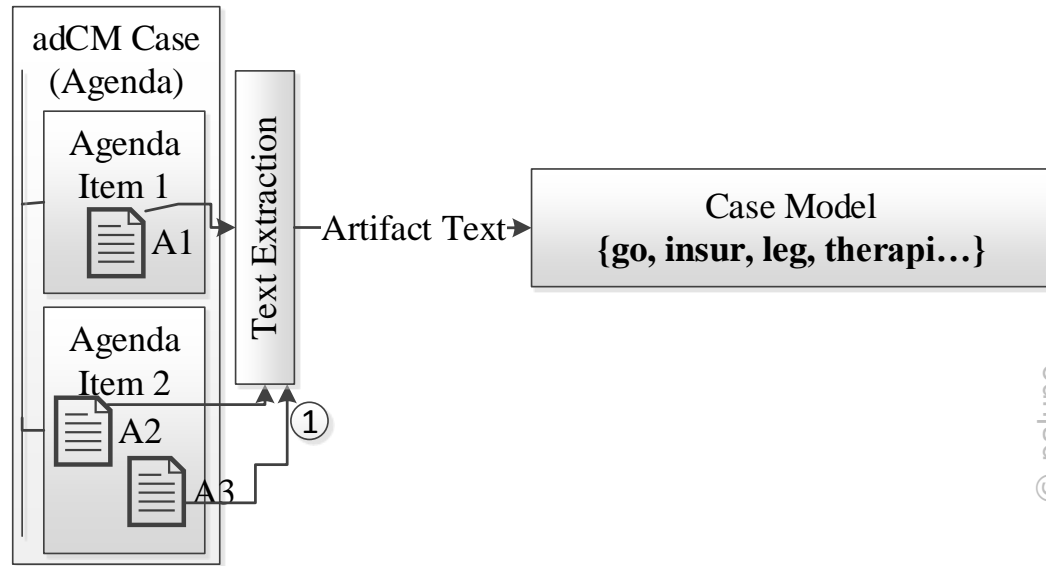


Evaluation

Ontology-based Model vs. Text-based Model



VS.



Evaluation Scenarios & Metrics

■ Scenarios

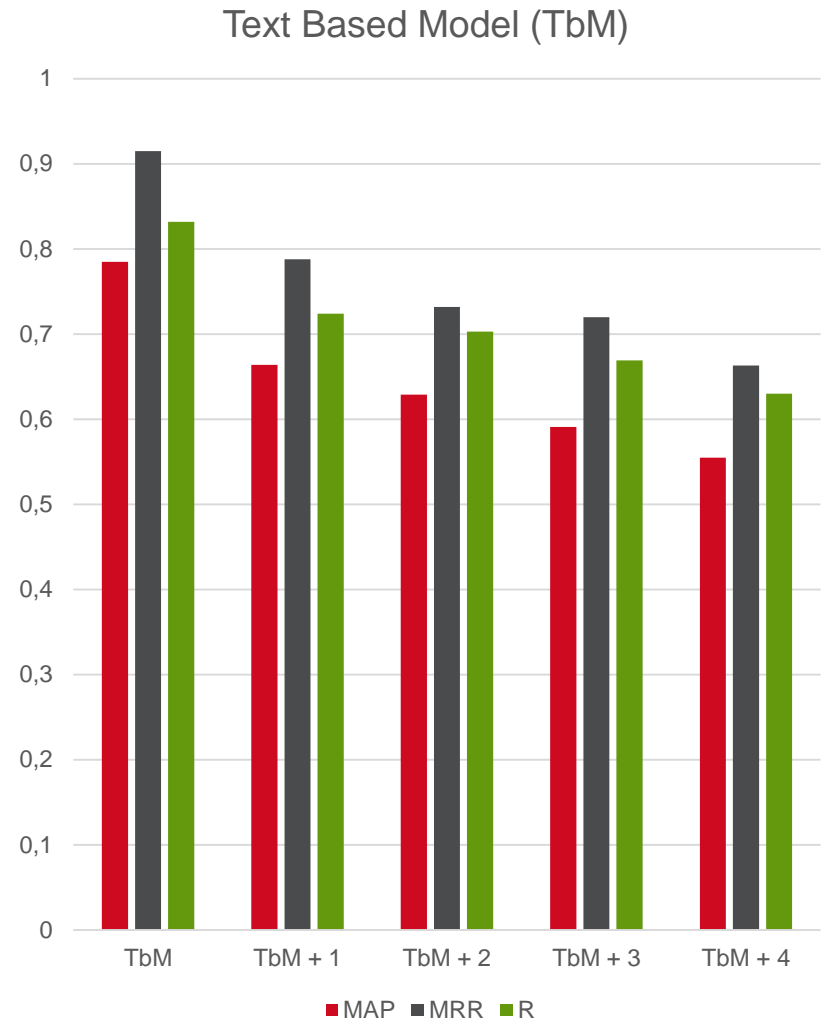
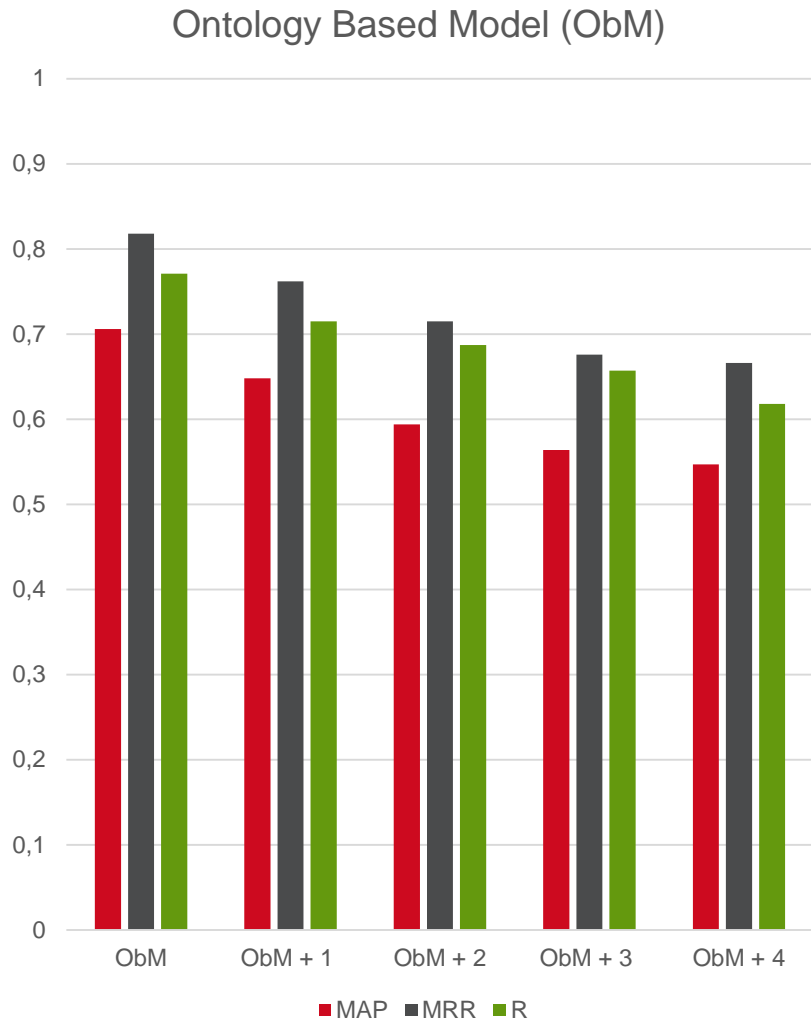
- **S1.** Cases built of homogenous information concerning only the particular medical problem
- **S2.** Cases contain medical history of the patient
- **S3.** Cases contain heterogeneous information that is not only from the medical domain

■ Evaluation Metrics

- **Recall (R):** relevant hits found/relevant hits existing
- **Mean reciprocal rank (MRR):** $1/r$ where r is the position or rank of the first correct case
- **Mean average precision (MAP):** the fraction of the documents retrieved that are relevant to the user's information need

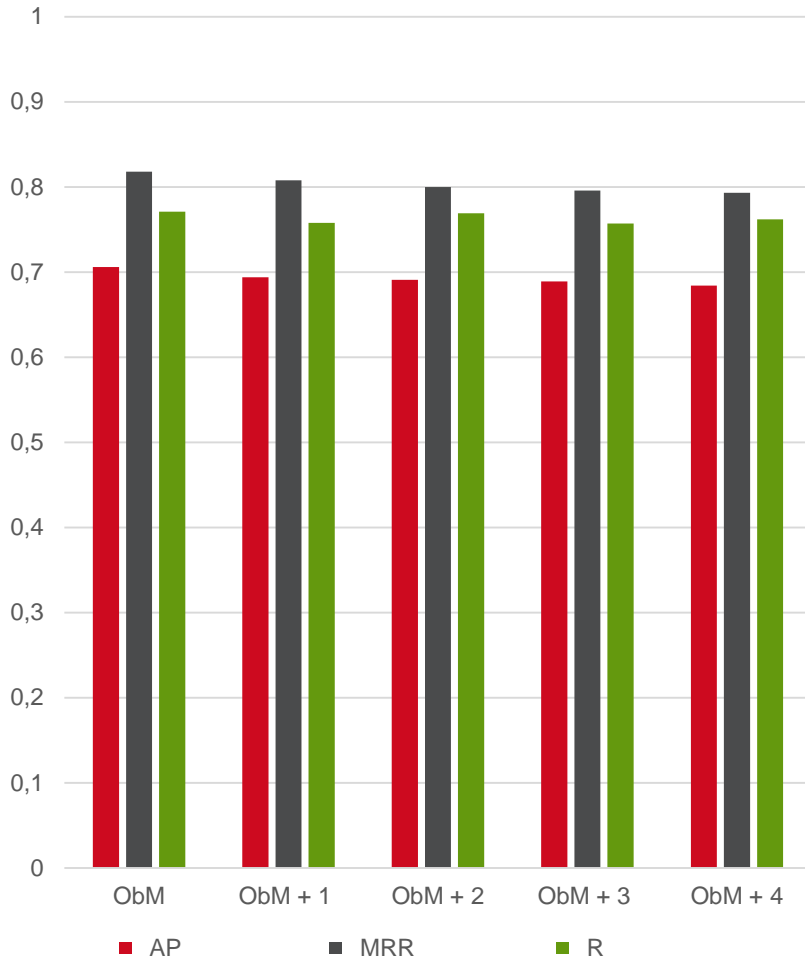
RESULTS

Scenario 1 and 2: Results

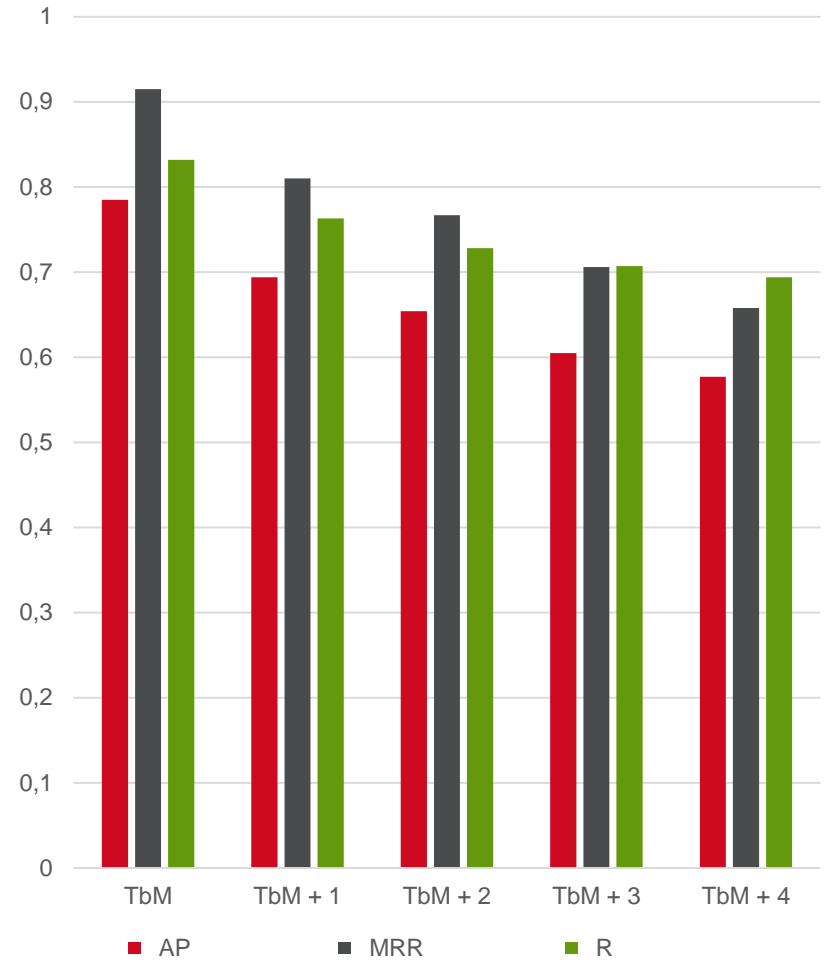


Scenario 1 and 3: Results

Ontology Based Model



Text Based Model



Conclusion

- Integration of the Information Recommender System (IRS) into the adCM tool
- A hybrid CBR recommender system
 - Ontology based retrieval
 - Text based prioritization and adjustment of artifacts
- Evaluation of ontological case model and textual model
 - Cases containing homogenous domain information are better handled **by textual case model**
 - Cases with information from heterogeneous domains are better handled by ontological model
 - Further evaluation required

Questions?

References

- [1] Aamodt, A. and Plaza, E. 1994. Case-based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Commun* 7, 1, 39–59.
- [4] Benner, M., Book, M., Brückmann, T., Gruhn, V., Richter, T., and Seyhan, S. 2012. Managing and Tracing the Traversal of Process Clouds with Templates, Agendas and Artifacts. In *Proceedings of the 10th International Conference on Business Process Management*. BPM'12. Springer-Verlag, Berlin, Heidelberg, 188–193. DOI=10.1007/978-3-642-32885-5_14.
- [5] BPTrends. 2009. *Case Management: Combining Knowledge With Process*. www.bptrends.com.
- [7] Davenport, T. H., Jarvenpaa, S. L., and Beers, M. C. 1996. Improving knowledge work processes. *Sloan management review* 37, 53–66.
- [10] Feldman, S. 2004. *The high cost of not finding information*. <http://www.kmworld.com/Articles/Editorial/Features/The-high-cost-of-not-finding-information-9534.aspx>. Accessed 15 November 2013.

Evaluation metrics

- Mean reciprocal rank (**MRR**): $1/r$ where r is the rank of the first correct result, or 0 if there are no correct results within the top 5 results
- Recall(**R**): $\frac{\text{relevant hits found}}{\text{relevant hits existing}}$
- Precision at n (**P_n**): (relevant hits within first n hits) / n .
- Average precision (**AP**): $\frac{1}{m} \times \sum_{k=1}^n (P(k) \times rel(k))$, where m is the number of relevant documents for an information need Q_j , k is the rank in the sequence of retrieved documents, n is the number of retrieved documents, $P(k)$ is the precision at cut-off k and $rel(k)$ is a function equaling 1 if the item at rank k is a relevant document, zero otherwise.
- Mean average precision (**MAP**): $\frac{1}{N} \sum_{j=1}^N AP(Q_j)$ where N is the number of queries in the test set