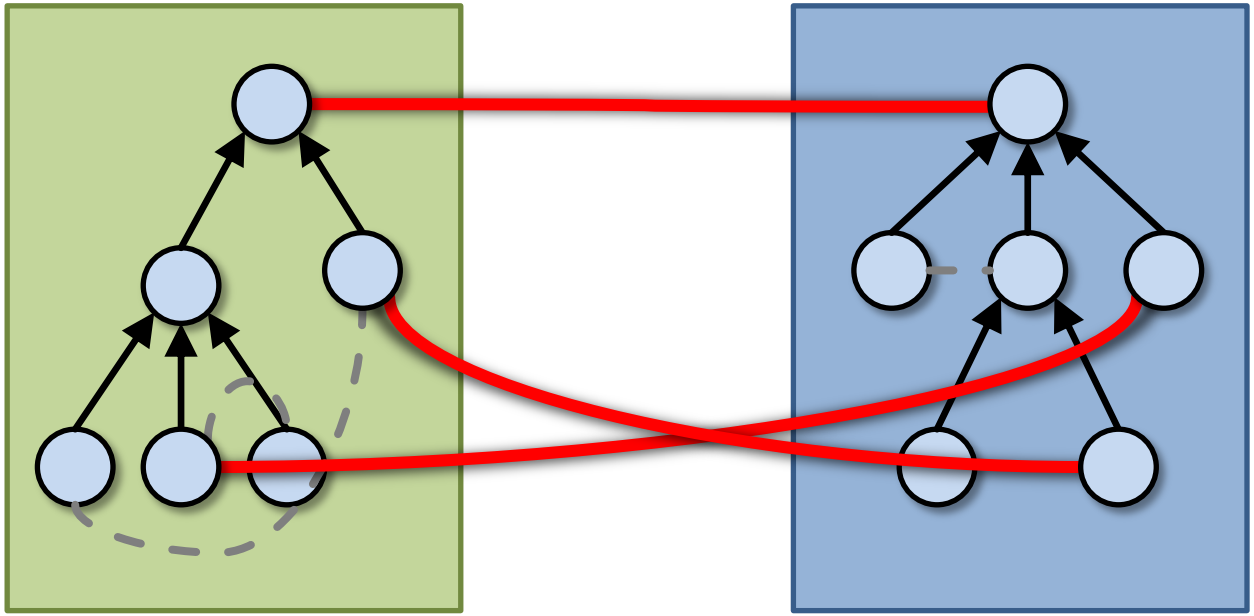
HPI Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

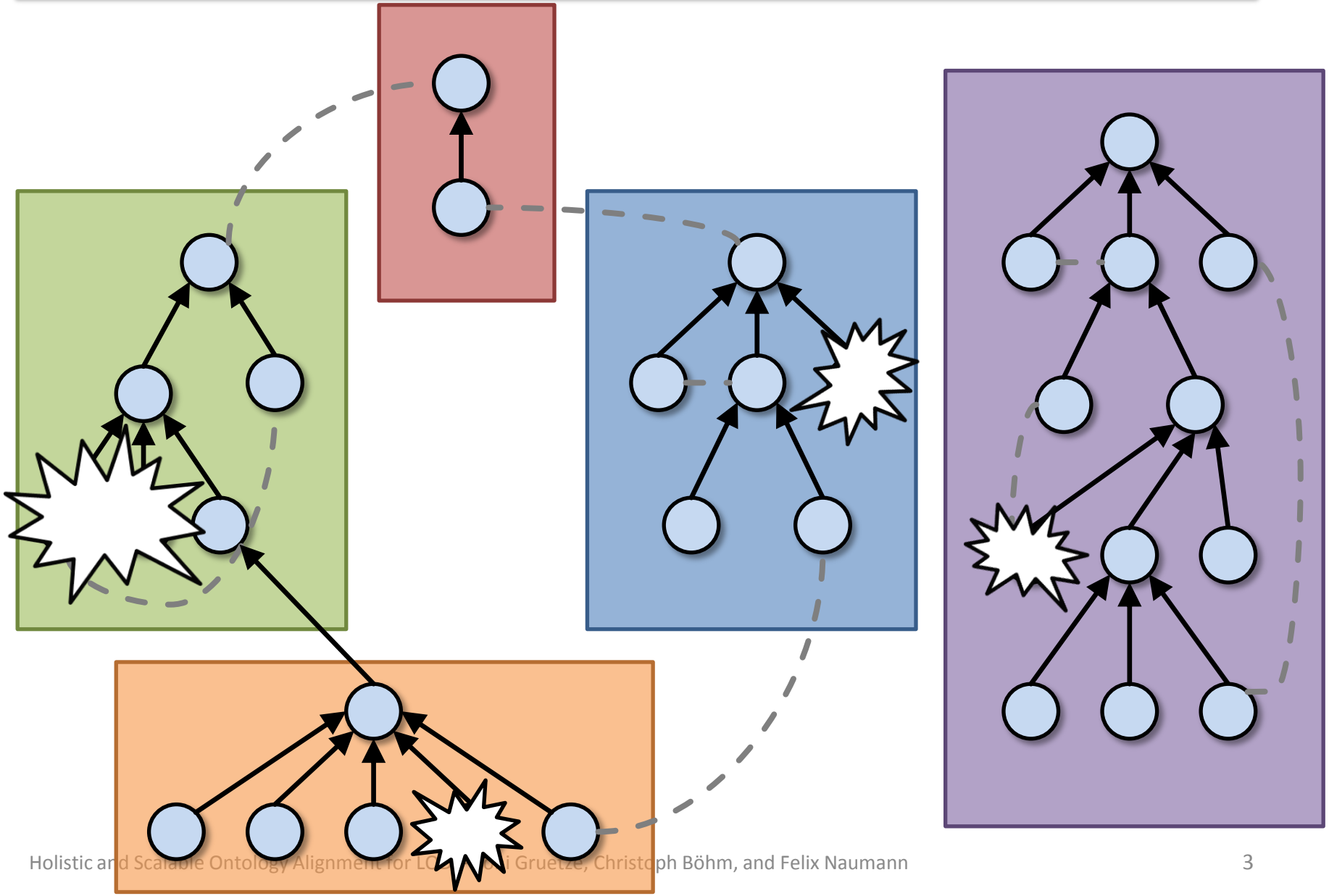
Holistic and Scalable Ontology Alignment for LOD

Toni Gruetze, Christoph Böhm,
and Felix Naumann

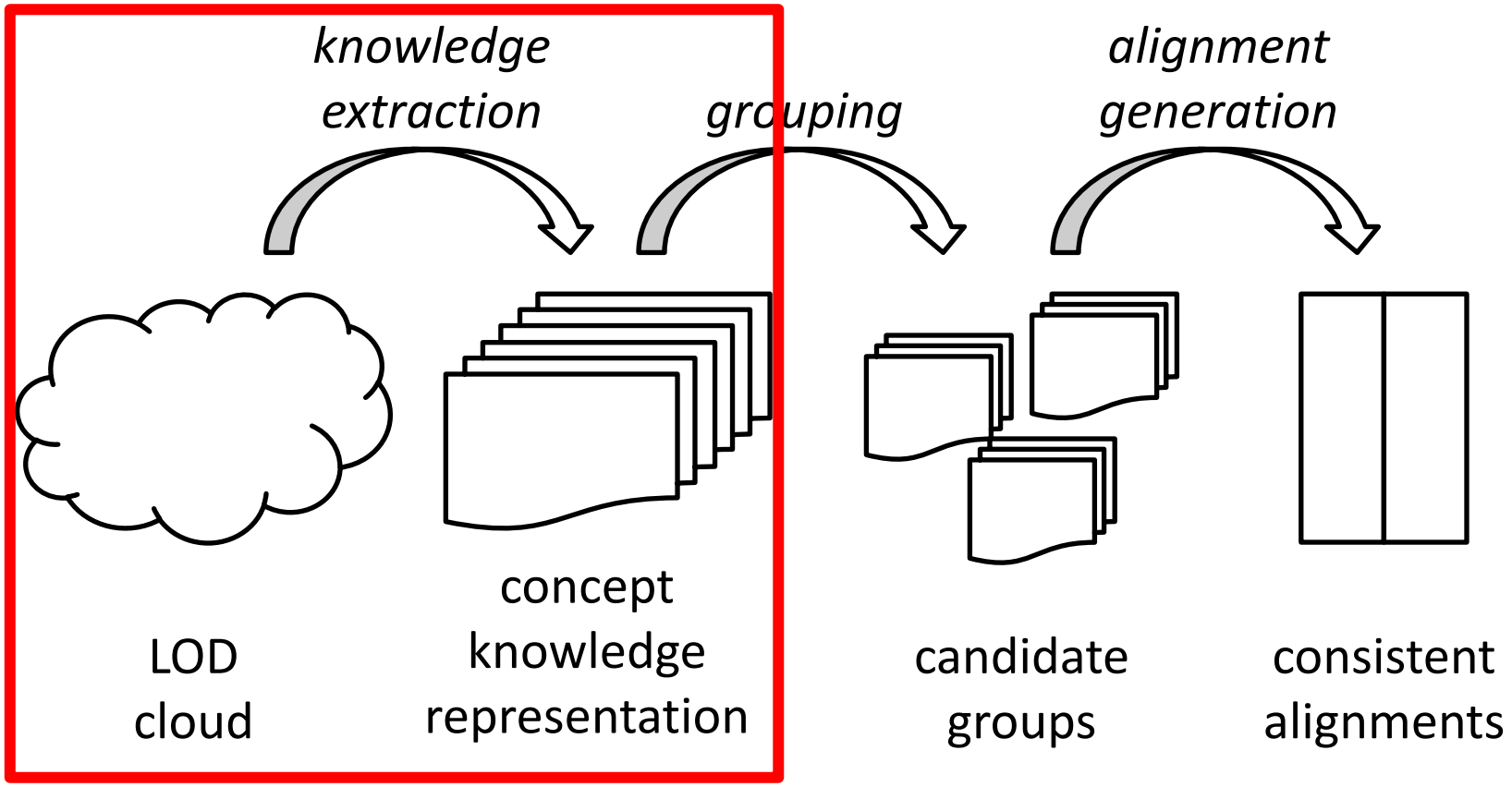
- Yet another Matching Algorithm?



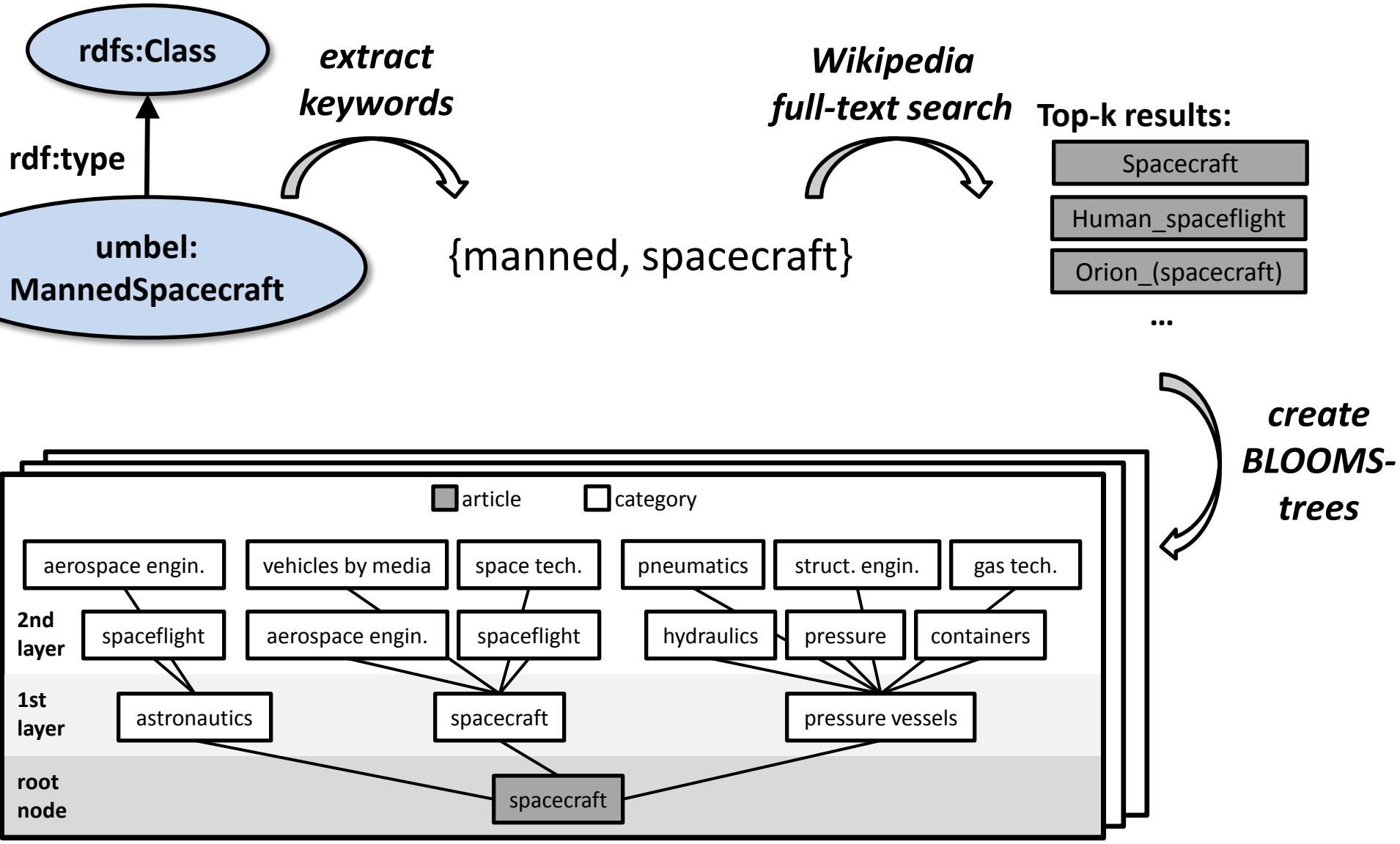
Holistic and Scalable Ontology Alignment for LOD



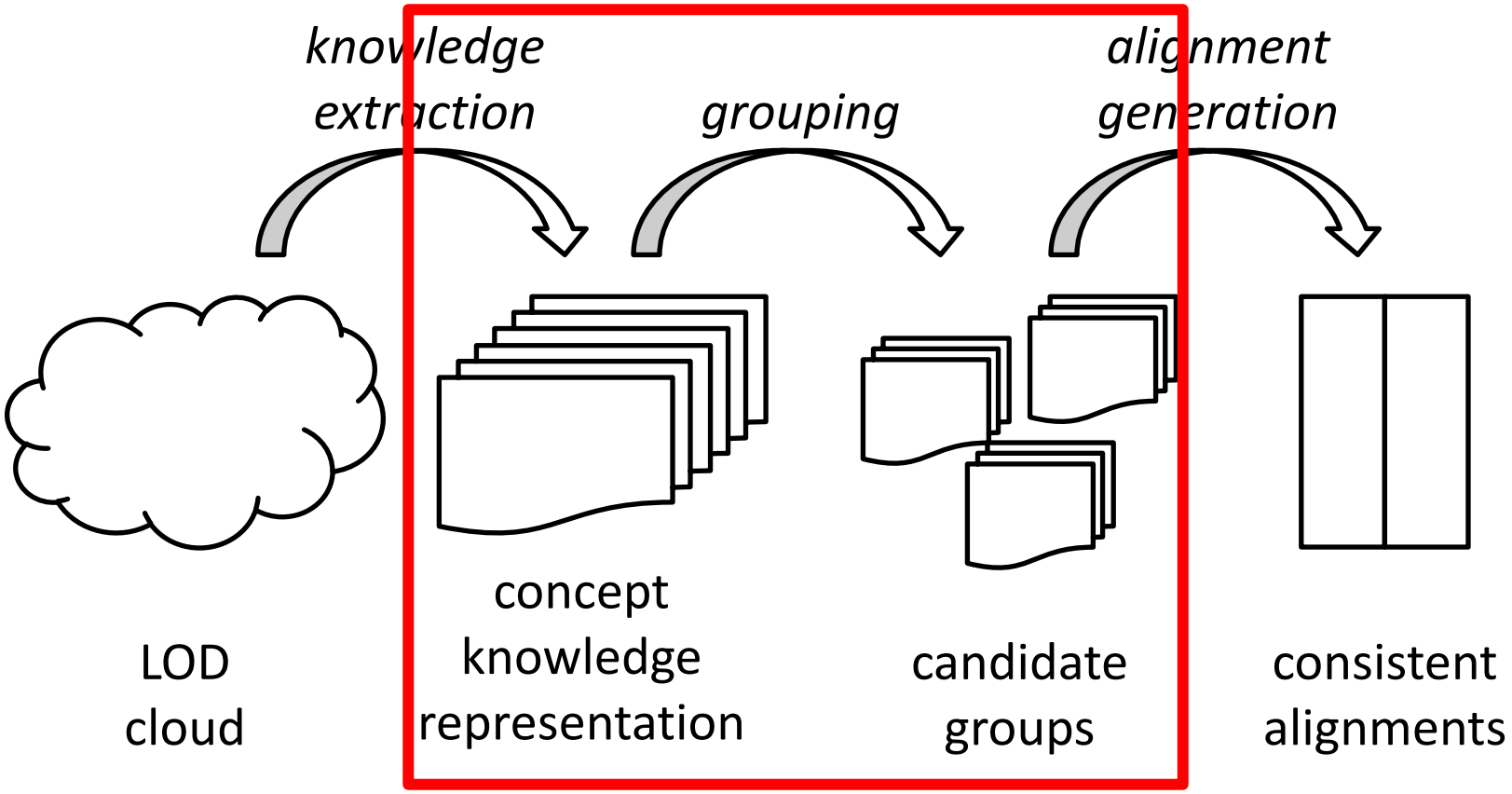
Approach – Overview



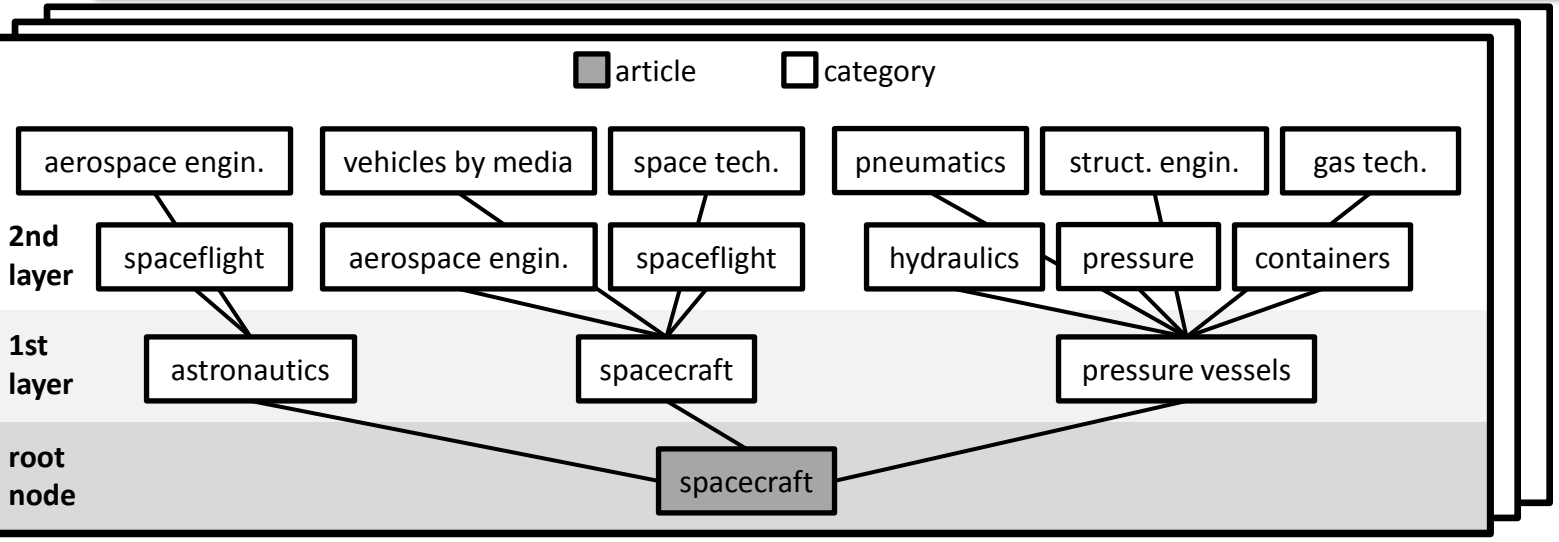
Knowledge Representation (BLOOMS[2,3])



Approach – Overview



Grouping

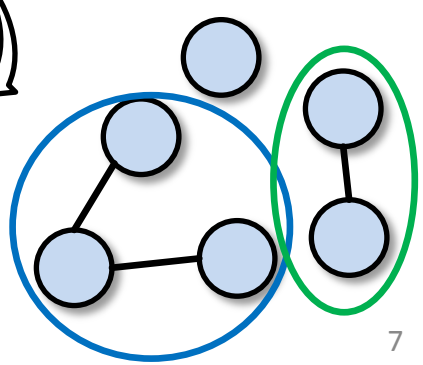


PPjoin[4]

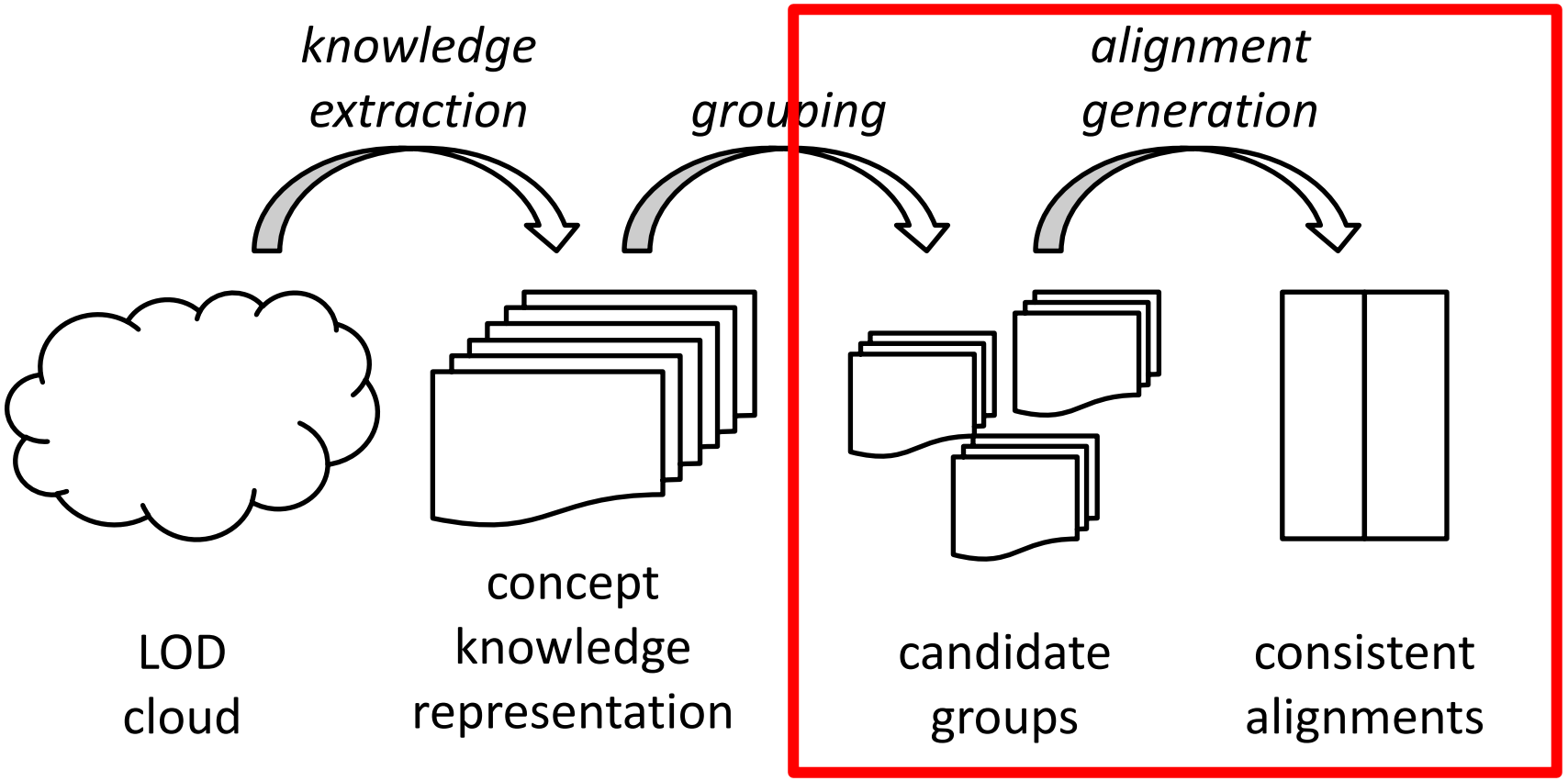
topic set extraction

{aerospace engin., spaceflight, spacecraft, human spaceflight}

set similarity join



Approach – Overview



Alignment Generation

1. Extend group by adding related forests
2. Compare all forest pairs
   - Based on BLOOMS+ tree overlap measure[3]
   - Extract candidate matches with high similarities
3. Create an alignment graph
   - Iteratively add candidate matches with highest similarity
   - Check for semantic conflicts → ASMOV[5]
   - Infer further necessary alignments
4. Extract alignments from the alignment graph

Experiments

- Billion Triple Challenge 2011 Dataset[1]
- Hardware
  - Windows machine with Java 6
  - 8-cores à 2.66 GHz
  - 30GB RAM
- Manual annotation of a result sample with 3 classes:
  - Equivalent:

    yago:PsychoactiveFungi and umbel:HallucinogenicMushroom
  - Similar:

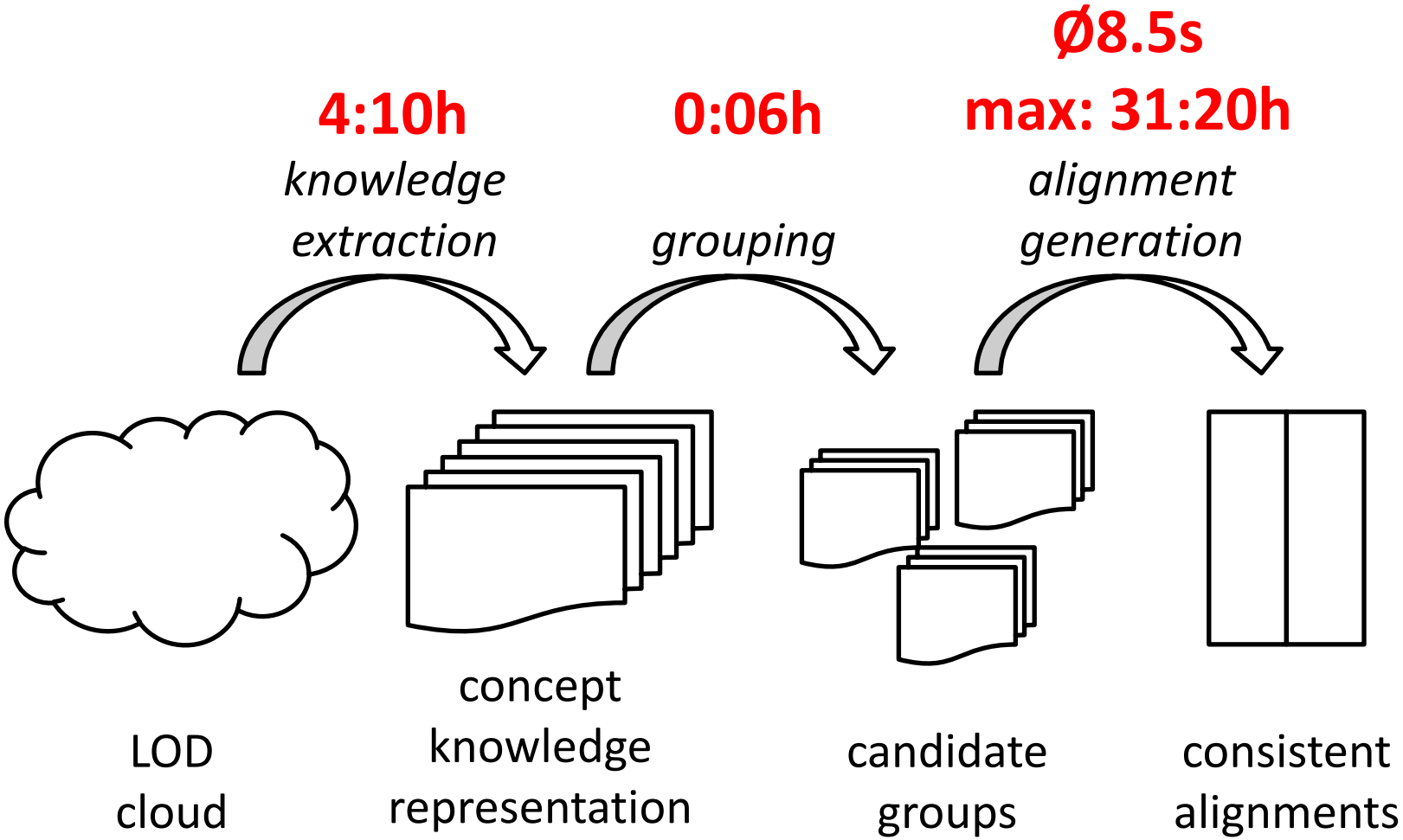
    daml:Ammters and umble:Voltmeter
  - Not equivalent:

    yago:Outlaws and umbel:MotorcycleClub

Evaluation – BTC'11: Results

$O(f_1, f_2)$	$P(eq)$	$P(eq \cup sim)$	align.
$O = 1.0$	0.964 ± 0.036	0.964 ± 0.036	601
$1.0 > O \geq 0.95$	0.575 ± 0.143	0.725 ± 0.129	133 317
$0.95 > O \geq 0.9$	0.481 ± 0.145	0.706 ± 0.131	72 131
$0.9 > O \geq 0.85$	0.071 ± 0.066	0.294 ± 0.131	6 921
$0.85 > O \geq 0.8$	0.053 ± 0.053	0.200 ± 0.114	4 279
$0.8 > O \geq 0.75$	0.053 ± 0.053	0.126 ± 0.092	3 139
$0.75 > O \geq 0.7$	0.071 ± 0.066	0.331 ± 0.136	1 950

Evaluation – BTC‘11: Runtime



Conclusion

- Graph data matching problem
- Abstract process
- Implementation of the process using a combination of available methods, namely:
  - BLOOMS[2,3]
  - PPjoin[4]
  - ASMOV[5]
- Evaluation on BTC‘11 shows good results

References

1. http://km.aifb.kit.edu/projects/btc-2011/
2. Ontology Alignment for Linked Open Data.
   Jain, P., Hitzler, P., Sheth, A. P., Verma, K., & Yeh, P. Z.; ISWC2010
3. Contextual Ontology Alignment of LOD with an Upper Ontology: A Case Study with Proton.
   Jain, P., Yeh, P. Z., Verma, K., Vasquez, R. G., Damova, M., Hitzler, P., & Sheth, A. P.; ESWC2010
4. Efficient similarity joins for near duplicate detection.
   Xiao, C., Wang, W., Lin, X., & Yu, J. X.; WWW2008.
5. Ontology matching with semantic verification.
   Jean-Mary, Y. R., Shironoshita, E. P., & Kabuka, M. R.; Web Semantics 2009