Knowledge Extraction from Description Logic Ontologies and the Applications in Industries

Jieying Chen

SIRIUS, University of Oslo, Norway

jieyingc@ifi.uio.no

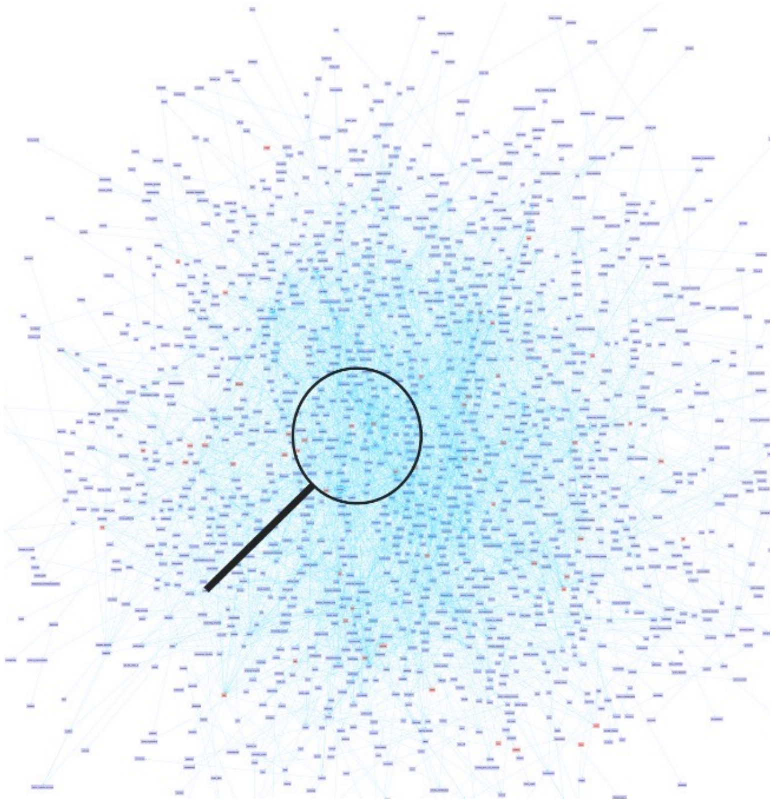
Background

- Ontologies are used to
  - represent knowledge and data
  - do reasoning tasks, like query answering

- Ontologies are being *continuously* developed and extended.

- Examples of biomedical ontologies:
  - SNOMED CT : 340 000 axioms
  - NCBI : 850 000 axioms
  - MESH : 400 000 axioms
  - NCIt : 75 000 axioms

- Goal: compute a subontology that is most relevant to users



Approaches

- Justifications (ESWC22)

- Ontology Modularity: Subsumption / Deductive Modules
  - ELH ontologies (ISWC17)
  - ELH^r ontologies (JELIA19)
  - ALCH ontologies (IJCAI20)

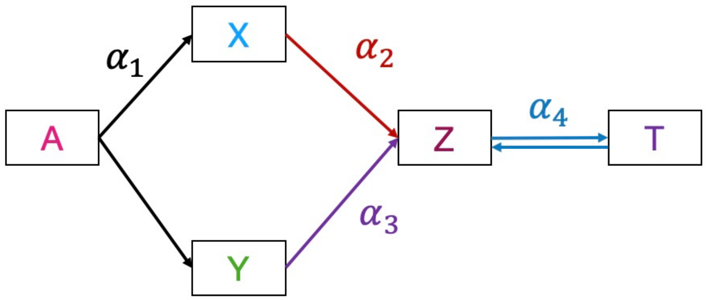
- Best Excerpt (ISWC17)

Justifications

Definition (Justification)

A justification for $\mathcal{O} \models \alpha$ is a subset $\mathcal{M} \subseteq \mathcal{O}$ such that $\mathcal{M} \models \alpha$ and for any $\mathcal{M}' \subsetneq \mathcal{M}$, $\mathcal{M}' \not\models \alpha$.

α_1	$A \sqsubseteq X \sqcap Y$
α_2	$X \sqsubseteq Z$
α_3	$Y \sqsubseteq Z$
α_4	$Z \equiv T$



JUST($A \sqsubseteq Z$)=$\{\{\alpha_1, \alpha_2\}, \{\alpha_1, \alpha_3\}\}$

Justifications

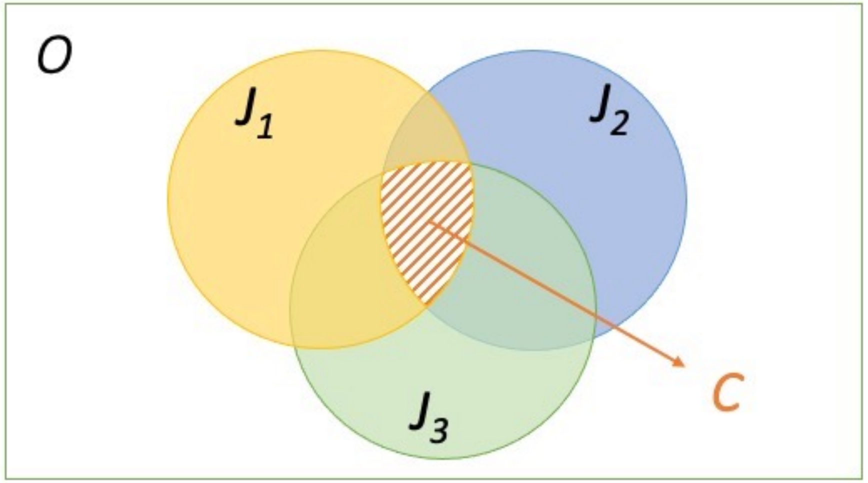
Definition (Justification)

A justification for $\mathcal{O} \models \alpha$ is a subset $\mathcal{M} \subseteq \mathcal{O}$ such that $\mathcal{M} \models \alpha$ and for any $\mathcal{M}' \subsetneq \mathcal{M}$, $\mathcal{M}' \not\models \alpha$.

- Applications:
  - diagnose ontologies
  - repair inconsistent ontologies
- There might exist exponentially many justifications!
- It is difficult to compute all justifications for expressive Description Logics.

Approximation of all Justifications

- Lower-bound: intersection of all justifications (C)



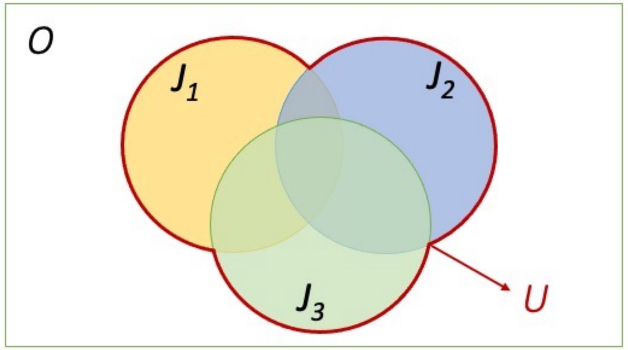
$$C = \{\beta \in O \mid O \setminus \{\beta\} \nvDash \alpha\}$$

Our Algorithm runs in **P** time!

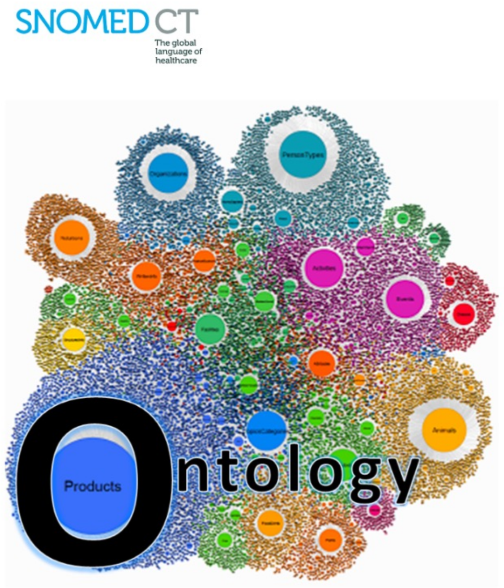
Approximation of all Justifications

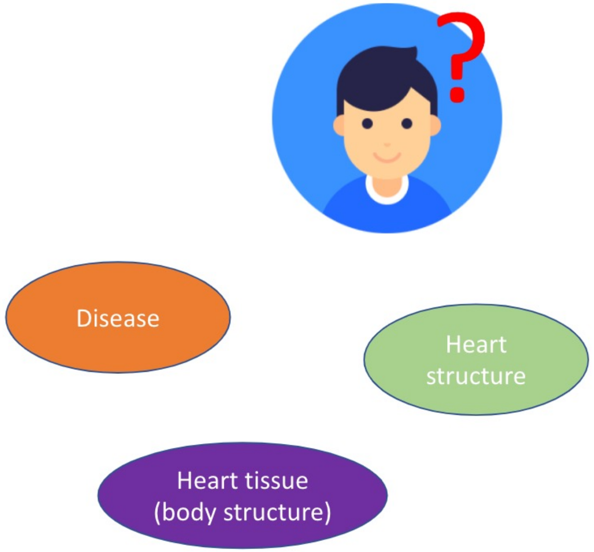
Upper-bound: union of all justifications (U)

- Naïve approach
- Black-box approach
  - inspired by the algorithm of computing all minimal hitting sets [Reiter, JAI1987]
  - OWL 2 ontologies
- MUS-Membership approach
  - Idea: check each axiom whether it is a member of some justification
  - MUS-MEM encoding:
    - Transfer all axioms in O and the given inclusion α to CNF formulae
    - For each CNF formula, check whether it is a member of MUS using SAT-solver
  - Works on ALC ontologies



Motivation: ontology modularity





Signature Σ = {Disease, Heart structure, Heart tissue (body structure)}

Definition of Ontology Modules

Definition (Subsumption/Deductive Module)

A subset $\mathcal{M} \subseteq \mathcal{O}$ is called a deductive module of $\mathcal{O}$ w.r.t. Σ iff $\mathcal{M} \models \alpha$ for every $\mathcal{O} \models \alpha$, $\mathsf{sig}(\alpha) \subseteq \Sigma$.

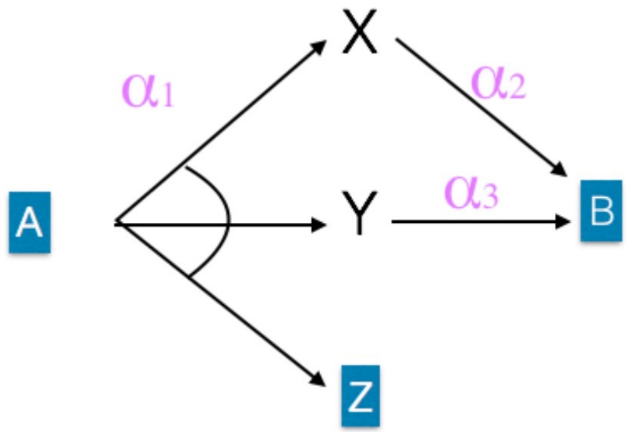
A deductive module is a subset of an ontology that can preserve subsumption queries.

- $\mathcal{M}$ is called a minimal module iff $\mathcal{M}$ is minimal w.r.t. $\subsetneq$.
- exponentially many in the worst case

Example of Ontology Modules

Example

Let $\mathcal{T} = \{\ \alpha_1 := A \sqsubseteq X \sqcap Y \sqcap Z, \quad \alpha_2 := X \sqsubseteq B, \quad \alpha_3 := Y \sqsubseteq B\ \}$
and $\Sigma = \{A, B, Z\}$.



- Min-Mod$(\mathcal{T}, \Sigma) = \{\{\alpha_1, \alpha_2\}, \{\alpha_1, \alpha_3\}\}$

Algorithms for Computing Minimal Deductive Modules

- ELH ontologies (conjunctions, existential restrictions, role inclusions)
  - Glass-box algorithm (ISWC17)
  - Black-box algorithm (GCAI18)
- ELH^r ontologies (ELH + range restriction)
  - Glass-box algorithm (JELIA19)
- ALCH ontologies (ELH + negation, universal restrictions)
  - Hybrid algorithm (IJCAI20)

Evaluation

- Our algorithms are practical in real-world ontologies: Snomed CT etc.
- Size: minimal modules are much smaller than other modules

TABLE – Modules of ontologies in the DL Classifications tract of OWL reasoner evaluation for 100 random signatures containing 60 concept names and all role names (min / max / med / stdev)

Size Min-Mod	0 / 1,179 / 179 / 119
Size AMEX-Mod	0 / 2,333 / 258 / 142
Size Locality-Based-Mod	0 / 65,566 / 1,022 / 196

$|\text{Min-Mod}| \leq |\text{AMEX-Mod}| \leq |\text{Locality-Based-Mod}|$

Motivation of Ontology Excerpts

- No control of module size
- The size of module can be quite large

Definition (k-Excerpt)

A k-excerpt of $\mathcal{T}$ w.r.t. Σ is a subset $\mathcal{E}$ of $\mathcal{T}$ such that

$$|\mathcal{E}| \leq k.$$

Definition of Best k-Excerpts

Definition (Best k-Excerpt)

A best k-excerpt of $\mathcal{T}$ w.r.t. Σ under μ is a k-element subset $\mathcal{E}$ of $\mathcal{T}$ such that

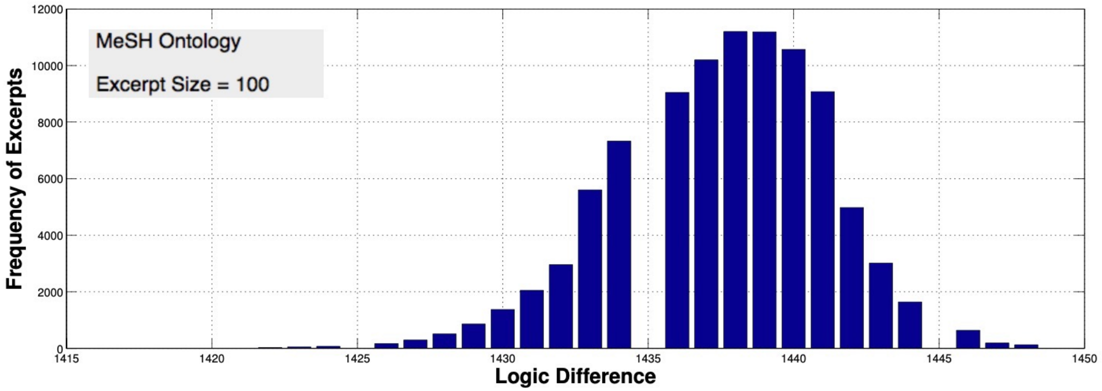
$$\mu(\mathcal{T}, \Sigma, \mathcal{E}) = \min\{\mu(\mathcal{T}, \Sigma, \mathcal{E}') \mid \mathcal{E}' \text{ is a } k\text{-excerpt of } \mathcal{T}\}.$$

- Measure quality of excerpts using Logical Difference

$$\mu(\mathcal{T}, \Sigma, \mathcal{E}) = \mid \mathsf{cWtn}_{\Sigma}(\mathcal{T}, \mathcal{E}) \mid$$

Do we need algorithms for Computing Best Excerpts?

✔ Excerpt quality varies !



more than 6.2×10^{180} excerpts !

- We need an algorithm to compute best excerpts directly.

Algorithm of Computing Best Excerpt

- Input: an ontology, a signature and k

Algorithm

- Compute $\langle X, \Sigma \rangle$-subsumption justifications for each concept name X in signature
- Encode best excerpt problem to partial Max-SAT problem
- Compute a solution using partial Max-SAT solver and transform the solution into best excerpts, e.g. Sat4j

Evaluation

TABLE – Times for computing best excerpts encoding the input for and running SAT4J

	Snomed CT				NCI			
Time (s)	Minimal	Maximal	Median	STDEV	Minimal	Maximal	Median	STDEV
MaxSat	0.01	1.27	0.15	0.35	0.01	0.16	0.01	0.02

Partial Max-SAT encoding is efficient !

Comparision: Justification vs. Minimal Module vs. Best Excerpt

- Common : a subset of ontology
- Differences :

	Best k-excerpts ($\mathcal{E}$)	Min. Mod. ($\mathcal{M}$)	Justifications
Input	$\mathcal{O}, \Sigma, k$	$\mathcal{O}, \Sigma$	$\mathcal{O}, \alpha$
Module Size	$\leq$ k	uncertain	
Answer to Query	unguaranteed	complete w.r.t. $\mathcal{O}, \Sigma$	complete w.r.t. $\mathcal{O}, \alpha$

- Connection :
  - $\mathcal{E} = \mathcal{M}_S$ when $k = |\mathcal{M}_S|$

    where $\mathcal{M}_S$ is the smallest minimal module, and $\mathcal{E}$ is a best k-excerpt of the ontology.

Industrial Projects

Digital Health

- Collaboration with Snomed International (KCAP19)

Energy

- Semantic projects with Aibel (with a master student)
- GeoDataPreparation with Equinor

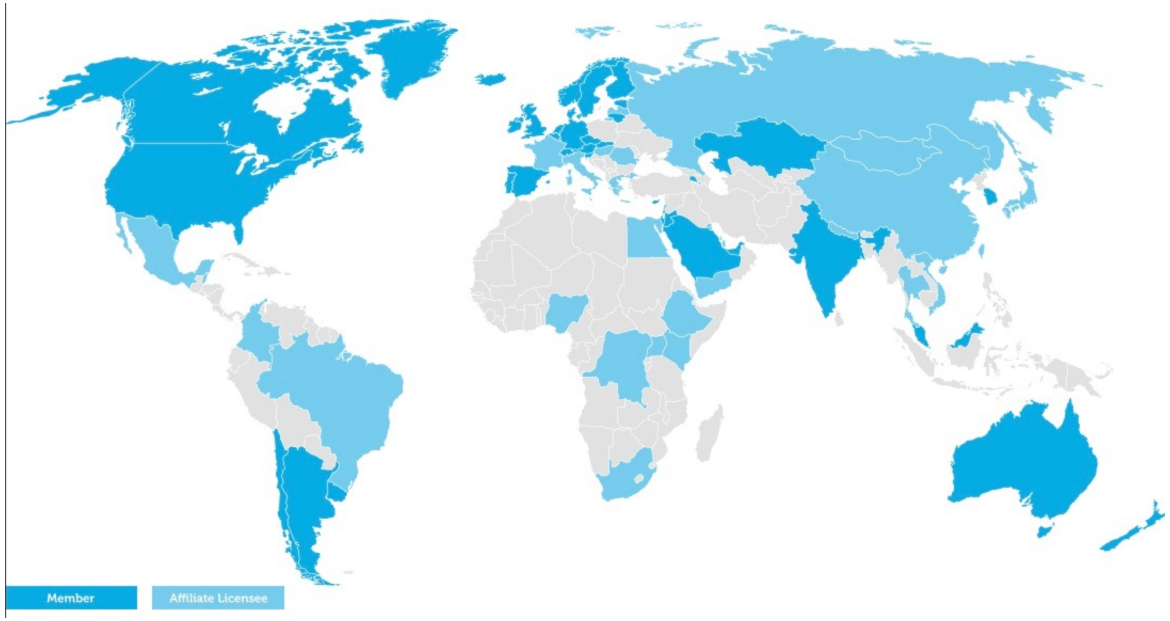
Manufacturing

- Combine semantic technologies with machine learning (IJCKG21)

Collaboration with Snomed International

Motivation

- 41 member countries and more than 5 000 individuals and organizations all over the world
- > 340 000 classes
- > 340 000 logical axioms
- Update and publish every six months



Ontology metrics:

Metrics

Axiom	1048856
Logical axiom count	349761
Declaration axioms count	349547
Class count	349428
Object property count	119
Data property count	0
Individual count	0
Annotation Property count	1

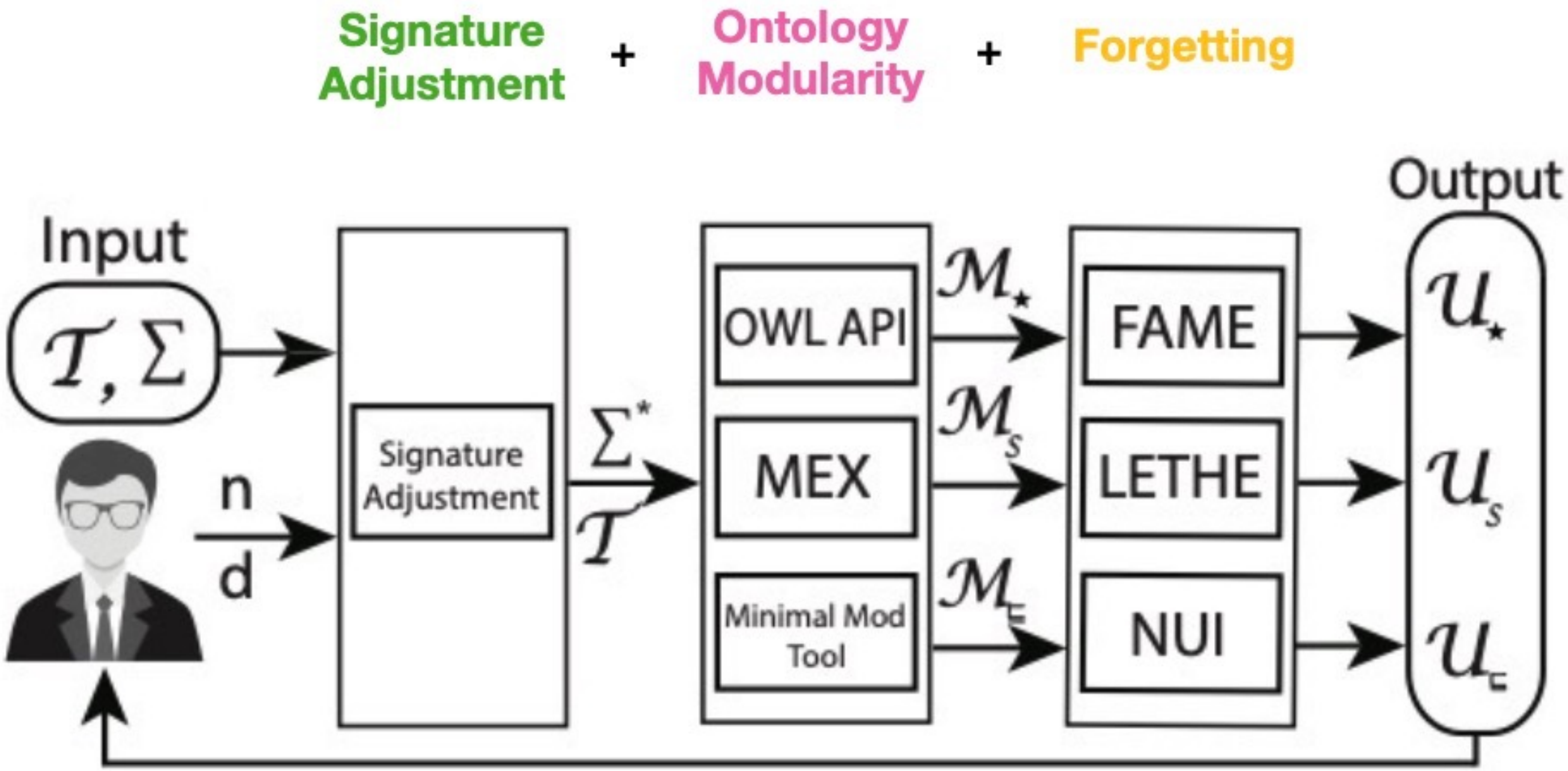
Class axioms

SubClassOf	243409
EquivalentClasses	106223
DisjointClasses	0
GCI count	20
Hidden GCI Count	184

Object property axioms

SubObjectPropertyOf	118
EquivalentObjectProperties	0
InverseObjectProperties	0

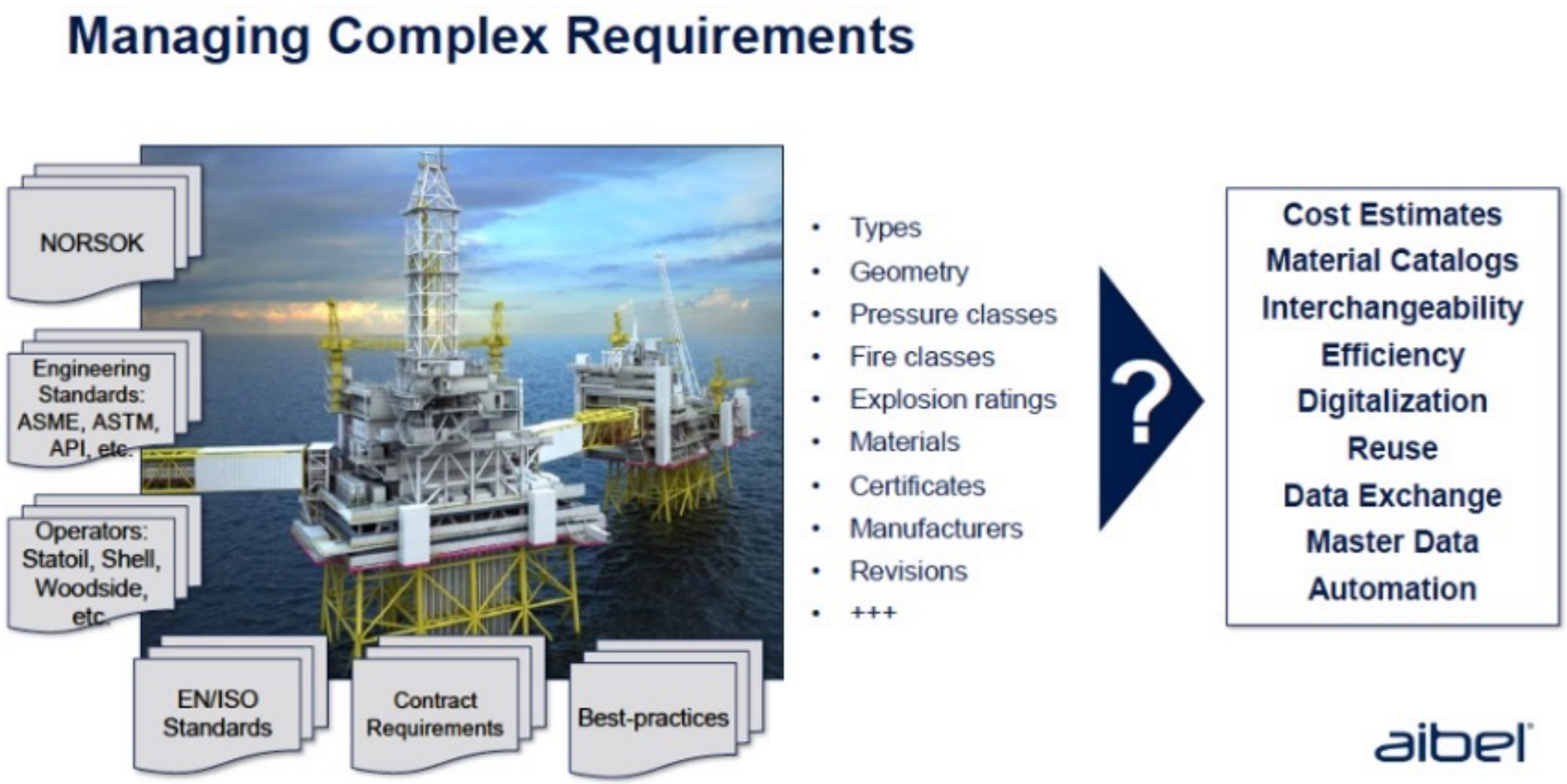
Workflow



K-CAP19

Collaboration with Aibel

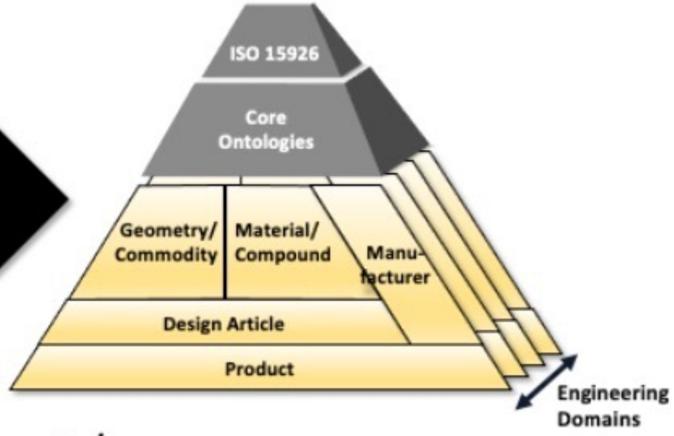
Motivation



Source: Christian M. Hansen (Aibel)

Material Master Data (MMD) for Piping Systems

Axioms	1.840.769
Logical axioms	535.512
Declaration axioms	106.674
Class count	98.133
Object property count	135
Data property count	723
Individual count	20.412
SubClassOf	505.376
EquivalentClasses	745
DisjointClasses	27
AnnotationAssertion	1.198.266



- Represent the contents of documents as an *ontology*
  - PDF documents: Engineering standards, client specifications
- Hierarchy of types and requirements for type membership
- Make explicit the *meaning* of document contents
  - Available to both humans and computers

Source: Christian M. Hansen (Aibel), David Cameron (SIRIUS)

Results

Processing time: 30 mins
-> less than 1 min

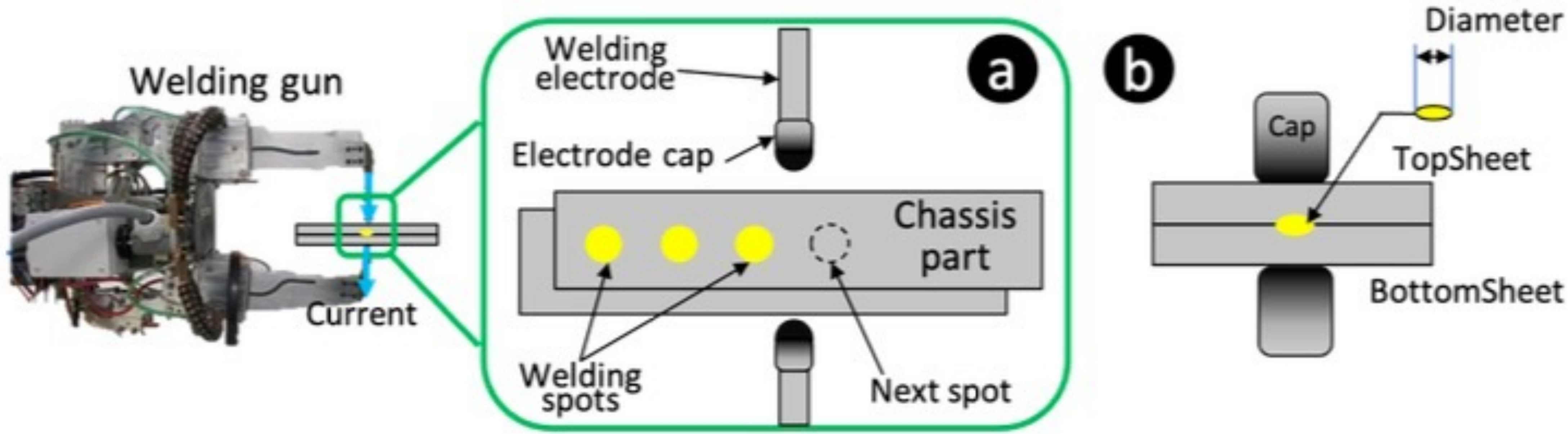
Query time: 30s ->2s

Precision: 100%

Size: 1.8 m -> 100k

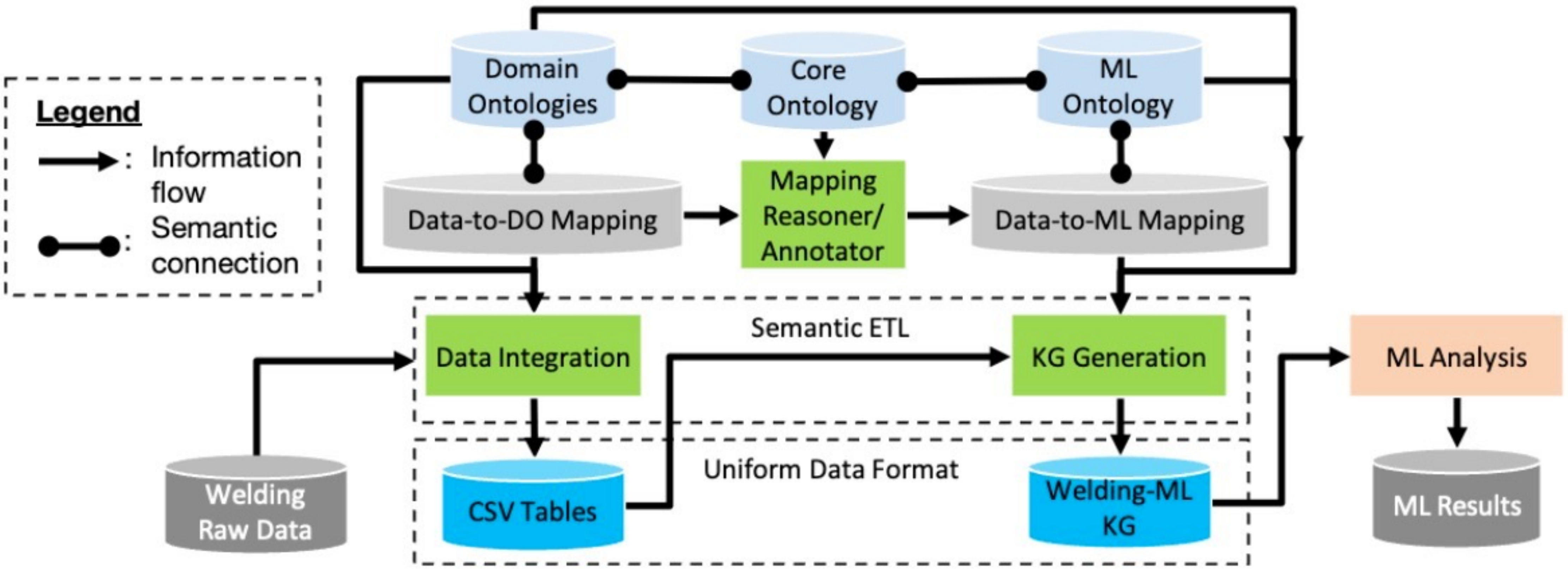
Collaboration with Bosch

User Case



Welding gun
Current
Welding electrode
Electrode cap
Chassis part
Welding spots
Next spot
a
b
Diameter
Cap
TopSheet
BottomSheet

Pipeline: generate KG for ML



IJCKG21

References

- *Jieying Chen*, Yue Ma, Rafael Peñaloza and Hui Yang: Union and Intersection of all Justifications. In Proceedings of the 19th European Semantic Web Conference (ESWC2022). To appear.
- Baifan Zhou, Dongzhuoran Zhou, *Jieying Chen*, Yulia Svetashova, Gong Cheng and Evgeny Kharlamov: **Scaling Usability of ML Analytics with Knowledge Graphs:Exemplified with A Bosch Welding Case.** In Proceedings of the 10th InternationalJoint Conference on Knowledge Graphs. To appear.
- Dongzhuoran Zhou, Baifan Zhou, *Jieying Chen*, Gong Cheng and Evgeny Kharlamov: **Towards Ontology Reshaping for KG Generation with User-in-the-Loop: Applied to Bosch Welding**. In Proceedings of the 10th International Joint Conference onKnowledge Graphs. To appear.
- Patrick Koopmann and *Jieying Chen*: **Deductive Module Extraction for Expressive Description Logics.** In Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI 2020), pp. 1636-1643, IJCAI Organization, 2020. [PDF][BibTeX]
- *Jieying Chen*, Ghadah Alghamdi, Renate A. Schmidt, Dirk Walther and Yongsheng Gao: **Ontology Extraction for Large Ontologies via Modularity and Forgetting**. In Proceedings of the 10th International Conference on Knowledge Capture (K-Cap 2019), pp. 45-52. ACM, 2019. [PDF][BibTeX]
- *Jieying Chen*, Michel Ludwig, Yue Ma and Dirk Walther: **Computing Minimal Projection Modules of Description Logic Terminologies**. In Proceedings of the 16th edition of the European Conference on Logics in Artificial Intelligence (JELIA 2019), vol.11468, pp.355-370. Springer, 2019. [PDF][BibTeX]
- *Jieying Chen*, Michel Ludwig, Dirk Walther: **Computing Minimal Subsumption Modules of Ontologies.** In Proceedings of the 4th Global Conference on Artificial Intelligence (GCAI 2018), Luxembourg, September 2018. EasyChair, 2018. [PDF][BibTeX]
- *Jieying Chen*, Michel Ludwig, Yue Ma and Dirk Walther: **Zooming in on Ontologies: Minimal Modules and Best Excerpts**. In Proceedings of the 16th International Semantic Web Conference (ISWC 2017), Part I. pp. 173–189. Springer, 2017. [PDF][BibTeX]
- *Jieying Chen*, Michel Ludwig, Yue Ma, and Dirk Walther: **Towards Extracting Ontology Excerpts.** In Proceedings of 8th International Conference on Knowledge Science, Engineering and Management (KSEM 2015), 2015, Proceedings. Lecture Notes in Computer Science, vol. 9403, pp. 78–89. Springer, 2015. [PDF][BibTeX]


thank you
danke
謝謝
gracias
merci
obrigado
ngiyabonga
teşekkür ederim
спасибо
dziękuję
go raibh maith agat
sukriya
kop khun krap
mochchakkeram
dakujem
grazie
arigatô
terima kasih
감사합니다
мерси
bedankt
tapadh leat
dank je
hvala
sagolun
takk
taiku


WHAT?
WHY?
WHEN?
WHERE?
HOW?
WHO?
WHICH?
WHOSE?