Towards Understanding Cognitive Aspects of Configuration Knowledge Formalization

A. Felfernig, S. Reiterer, M. Stettinger, and J. Tiihonen

alexander.felfernig@ist.tugraz.at

Paper source: http://dl.acm.org/citation.cfm?id=2701327

Overview

- Motivation
- Related Work
- Initial User Study
- Results
- Future Work
- Conclusions

Motivation

- Model-based Knowledge Representations
- Graphical Development Environments
- Automated Testing & Debugging
- Our Goal: Develop a more in-depth understanding of cognitive issues in knowledge base development (organization, representation, formalization)

Constraint Satisfaction Problem (CSP)

Definition (Constraint Satisfaction Problem and Solution – CSP). A CSP is a triple (V, D, C) where V represents a set of variables (V=$\{v_1, v_2, ..., v_n\}$), D describes the corresponding variable domains (D=$\{dom(v_1), dom(v_2), ..., dom(v_n)\}$), and C is a set of constraints (C=$\{c_1, c_2, ..., c_m\}$). A solution for a given CSP is represented by a complete set of variable assignments which is consistent with the set of constraints.

Example Configuration Model (as CSP)

$$V = \{wr, ip, rr\}$$
$$C = \{c_1, c_2, c_3, c_4, c_5\}$$
$$T = \{t_1, t_2, t_3, t_4\}$$
$$dom(wr) = \{low, medium, high\}$$
$$dom(ip) = \{shortterm, mediumterm, longterm\}$$
$$dom(rr) = \{3-6\%, 6-9\%, >9\%\}$$
$$c_1 : wr = medium \rightarrow ip \neq shortterm$$
$$c_2 : wr = high \rightarrow ip = longterm$$
$$c_3 : ip = longterm \rightarrow (rr = 3-6\% \vee rr = 6-9\%)$$
$$c_4 : rr = \; >9\% \rightarrow wr = high$$
$$c_5 : rr = 6-9\% \rightarrow (wr \neq low \wedge wr \neq medium)$$

Related Work: k-means Clustering of Constraints

$c_i \in C$	c_1	c_2	c_3	c_4	c_5	c_6	c_7
c_1	1.0	-	-	-	-	-	-
c_2	0.33	1.0	-	-	-	-	-
c_3	0.16	0.33	1.0	-	-	-	-
c_4	0.16	0.5	0.16	1.0	-	-	-
c_5	0.1	0.25	0.1	0.37	1.0	-	-
c_6	0.0	0.0	0.0	0.0	0.12	1.0	-
c_7	0.0	0.33	0.33	0.16	0.12	0.16	1.0

A. Felfernig, S. Reiterer, M. Stettinger, F. Reinfrank, M. Jeran and G. Ninaus, Recommender Systems for Configuration Knowledge Engineering, Workshop on Configuration, Vienna, Austria, CEUR Proceedings, pp. 51—54, 2013.

iteration	c_1	c_2	c_3	c_4	c_5	c_6	c_7
1	$1(cs)$	1	1	2	$2(cs)$	2	2
2	1	$1(cs)$	1	1	$2(cs)$	2	1

$$sim(c_a, c_b) = \frac{\sum_{v \in V} co\text{–}occurrence(v, c_a, c_b)}{|V|}$$

k groups generated by k-means clustering: each group has a centroid which is the constraint most similar to all others.

Related Work: Understandability and Constraint Grouping Criteria

Grouping approach	kba_1: *SOL*	kba_2: *CON*
Similar variables	21.43%	42.86%
Similar operators	30.77%	53.85%
Random	38.46%	76.92%

A. Felfernig, S. Reiterer, M. Stettinger, F. Reinfrank, M. Jeran and G. Ninaus, Recommender Systems for Configuration Knowledge Engineering, Workshop on Configuration, Vienna, Austria, CEUR Proceedings, pp. 51—54, 2013.

Table 5: Error rates for completing the tasks *find a solution (SOL)* and *find a conflict (CON)* depending on clustering approach (variable-based, operator-based, or random).

The best results were achieved with variable similarity-based clustering (N = 40 subjects).

Related Work: Understandability of Constraint Formalizations

kbb_1: SOL	errors	kbb_2: SOL	errors
$X \rightarrow Y$	21.43%	$X \rightarrow \neg Y$	14.29%
$\neg X \vee Y$	50.0%	$\neg X \vee \neg Y$	34.62%
$\neg Y \rightarrow \neg X$	96.43%	$Y \rightarrow \neg X$	50.0%
$\neg(X \wedge \neg Y)$	73.08%	$\neg(X \wedge Y)$	42.31%
$Y \leftarrow X$	25.0%	$\neg Y \leftarrow X$	16.67%

A. Felfernig, S. Reiterer, M. Stettinger, F. Reinfrank, M. Jeran and G. Ninaus, Recommender Systems for Configuration Knowledge Engineering, Workshop on Configuration, Vienna, Austria, CEUR Proceedings, pp. 51—54, 2013.

- Knowledge bases with different representations of „requires“ and „incompatibility“ relations.
- “Preferred” representations can serve as a basis for recommending constraint refactorings.

User Study

- Goal: develop an understanding of the impact of textual domain specifications on corresponding formalizations.
- Focus: Compatibility & Requires Constraints.
- Study Participants (subjects): 10 subjects, completed studies of Computer Science (2 Universities), 2-5 years of experience in developing constraint-based applications, 50% experiences from industry projects.
- Subjects with industrial background: two companies focusing on configuration and recommendation.

Example Task: Formalization of Requires Constraints

Natural Language Statement (s)	*Appropriate Formalization?*
A shortterm investment period as well as a high expected return rate require a high willingness to take risks.	1) ip=shortterm $\rightarrow$ wr = high. rr=high $\rightarrow$ wr = high.
	2) ip=shortterm $\vee$ rr=high $\rightarrow$ wr = high.
	3) (ip = shortterm $\rightarrow$ wr = high) $\vee$ (rr = high $\rightarrow$ wr = high).

☞100% correct formalizations.

Task: Redundancy Detection

Variables and Domains	*Redundant Constraints?*
$V = \{v_1, ..., v_{10}\}$ $dom(v_i = \{1,2,3\})$	$c_0 : v_8 \neq v_9.$
	$c_1 : v_9 \neq v_{10}.$
	$c_2 : v_2 \neq v_9.$
	$c_3 : v_3 = 1.$
	$c_4 : v_{10} \neq v_8.$
	$c_5 : v_9 \neq v_8.$
	$c_6 : v_4 > v_2.$
	$c_7 : v_2 \neq v_{10}.$
	$c_8 : v_8 \geq v_7.$
	$c_9 : v_9 > v_8.$
	$c_{10} : v_2 \neq v_8.$

Redundant constraint $c_i \in C$: $C - \{c_i\} \models c_i$

☞ 40% faulty responses, i.e., redundancy detection is a complex task even for simple settings.

Task: Formalization of Compatibility Constraints

Natural Language Statement (s)	*Appropriate Formalization?*
A high expected return rate is only compatible with a high willingness to take risks.	1) $rr = high \leftrightarrow wr = high.$
	2) $rr = high \wedge wr = high.$
	3) $rr = wr.$
	4) $rr = high \rightarrow wr = high.$

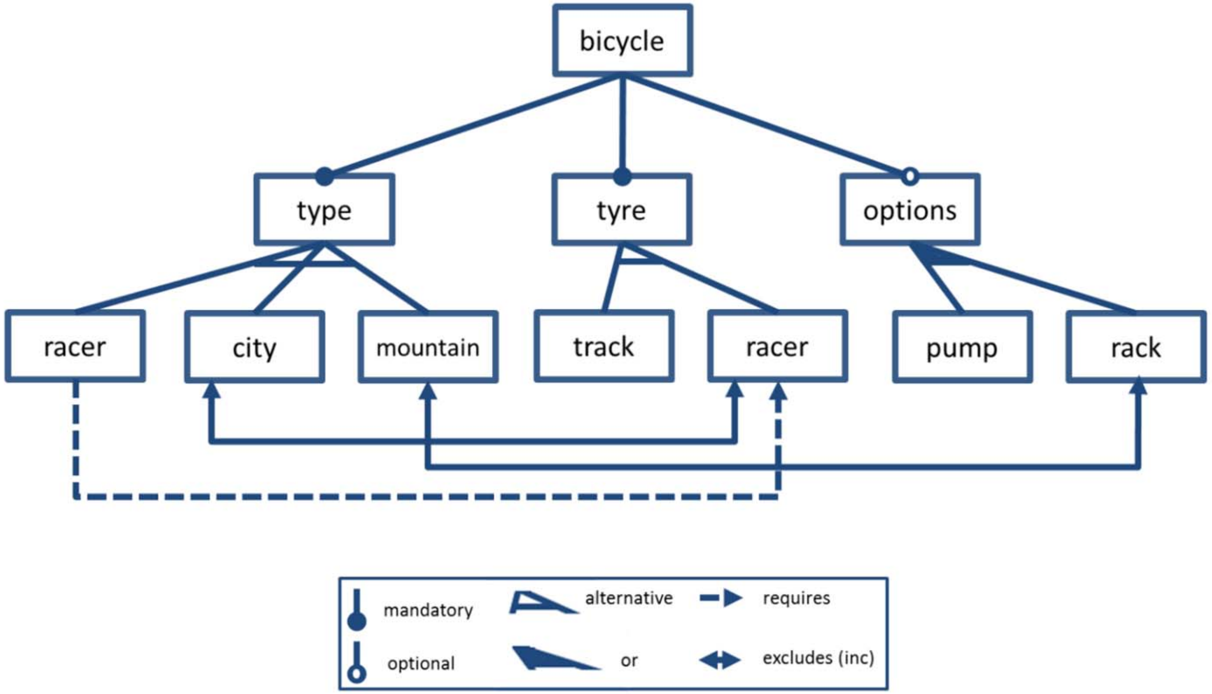
☞80% faulty formalizations. Potential assumption behind "only compatible": A with B and B with A.

Task: Formalization of Compatibility Constraints

Natural Language Statement (s)	*Appropriate Formalization?*
A shortterm investment period is only compatible with a low willingness to take risks (and vice-versa).	1) $ip = shortterm \leftrightarrow wr = low.$
	2) $ip = shortterm \land wr = low.$
	3) $ip = shortterm \rightarrow wr = low.$

☞90% correct formalizations due to a complete specification: both directions are taken into account.

Task: Model Description in Natural Language Text



☞Tendency of under-specifications of “or/alternative” relationships could be observed (90% of the cases).

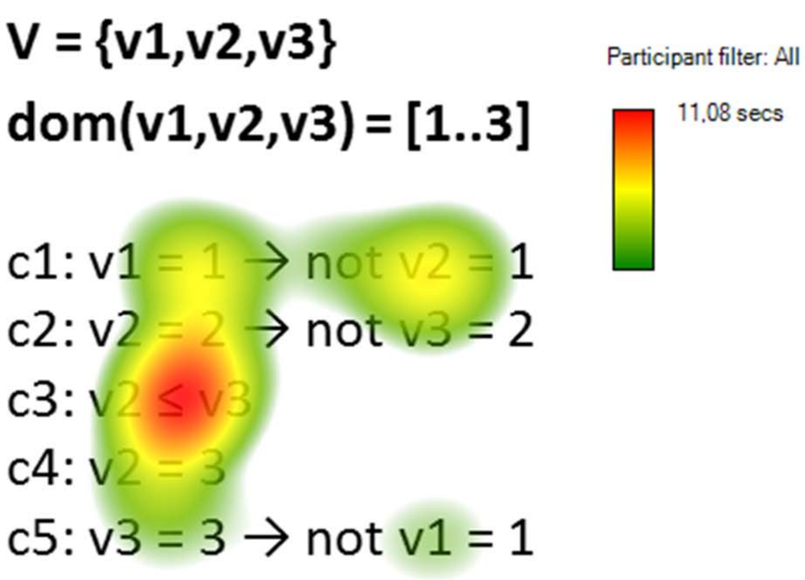
Preliminary Results (Summary)

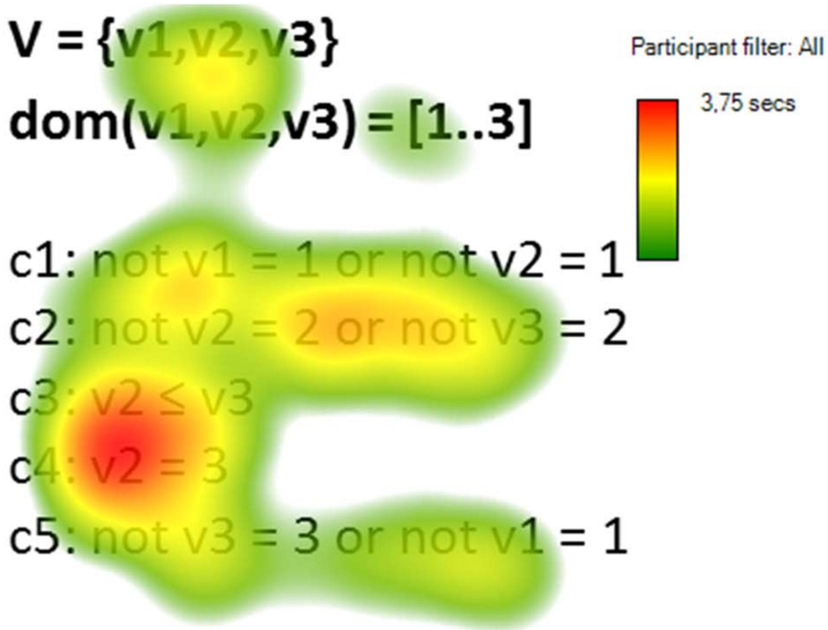
- Directed incompatibilities misunderstood („standard“ incompatibility assumed)
- Redundancy detection is a complex task even for low-complexity knowledge bases
- Similar results could be observed when study participants had to identify minimal diagnoses
- Underspecified „or/alternative“ semantics in domain descriptions
- „Direct“ translations (without further logical transformations) make constraints more maintainable

Future Work

- More in-depth analysis of industrial domain descriptions
- In-depth field study how domain experts specify variability knowledge (presented study limited)
- Empirical studies with larger user communities
- Extended set of knowledge types (e.g., generalization hierarchies, partonomies, resource constraints)
- Analyze in more detail existing work, for example, in Software Requirements Engineering

Future Work

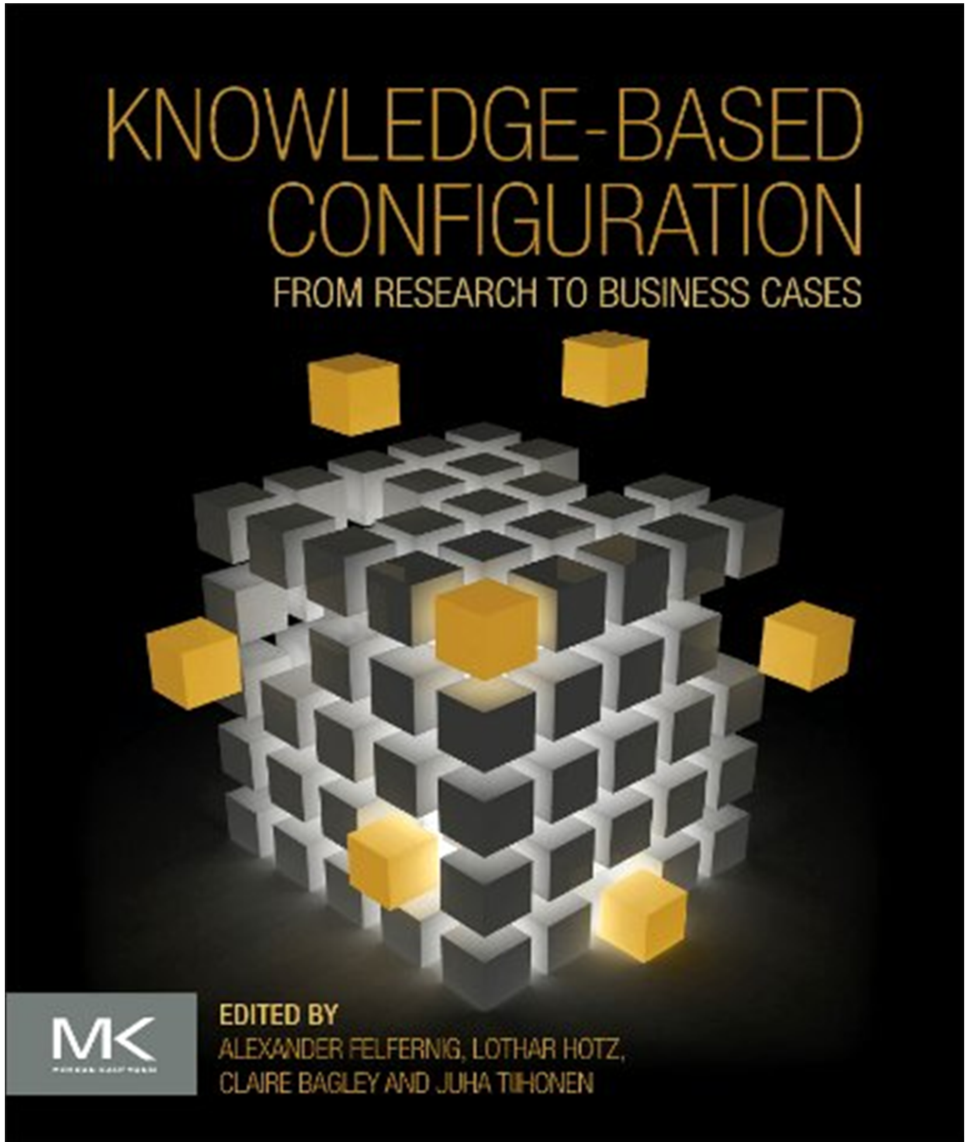




- Overview of areas, knowledge engineers looked at.
- Can be used, for example, for constraint ranking.

Conclusions

- Understanding cognitive processes in configuration knowledge engineering important
- Compact knowledge organization (e.g., clustering), understandable formalizations, no underspecified domain descriptions, …
- Focus of preliminary study: typical ways of formalizing domain knowledge
- Many open issues, for example, detailed analysis of industrial domain descriptions



Knowledge-Based Configuration by: *A. Felfernig*, *L. Hotz*, *C. Bagley*, and *J. Tiihonen*.

The purpose of this book is to expose the reader to a field of Artificial Intelligence that has been successfully integrated and used in the industry for more than 30 years. It provides configuration-related material for interested readers from the fields of industry, education, and research.

www.configurationbook.org

Thank You!