Using a cooperative agent system for a context-based discovery of interdatabase dependencies

(Extended Abstract)

M. Klusch
Institut für Informatik und Praktische Mathematik
Christian-Albrechts-Universität Kiel, Olshausenstr. 40, 24118 Kiel, Germany
E-Mail: *mkl@informatik.uni-kiel.d400.de*

Abstract

A novel approach towards the recognition of plausible interdatabase dependencies (IDD) using a federative cell system FCSI is presented. The architecture of the FCSI is designed as a system of co-operative intelligent autonomous agents (cells) each of them uniquely assigned to one local database system. The *FCSI aims for a cooperative solution for the problem of finding semantically related information while strictly respecting the autonomy requirements of each individual database system.* For this purpose first a *descriptive, terminological meta-knowledge and domain information representation* DIT *on top of the local conceptual database schema* is built at each agent by using user-specified intentional scripts. In addition, an respectively instantiated actual aspect world W enables a more flexible representation of *schema object aspects.* According to a prior known set of common terminological primitives cpc of respective local DITs remote *terminological classification* then serves as a basis for the *discovery of intentional IDD*s (i-IDD) between objects of different schemas with respect to some or all of their sofar related intentional aspects. Projections on respective associated *aspect valuations at schema and state level* isps determine *context-based data sharing* and result in restricted proposals for *interdatabase schema assertions* (IDSA). Methods for flexible utilitaristic *coalition building among the autonomous agents of the FCSI* are currently investigated in order to organize a *cooperative recognition process* of i-IDDs. In this paper the current status of ongoing research on the FCSI is reported.

1 Introduction

The approach of the FCSI is motivated by the idea of following the recently introduced paradigm of intelligent cooperative information systems ICIS [22] for solving one essential problem in the research area of interoperability of database systems [23]: *a context-based recognition of plausible interdatabase dependencies (IDD) between heterogeneous, autonomous database systems.*

An IDD describes the relationship between related data of different database schemas by an integrity constraint [9]. Any approach of declarative or functional specification of IDDs like in [28],[7],[9] presumes a-priori knowledge about where to find semantically related data as well as its availability. The related problem of tackling semantic heterogeneity[26] is aggravated by the need to respect in particular the association autonomy[27] of all respective database systems.

Recent criticism[5][21] on federated database systems FDBS[27] state in particular their lack of support with respect to this *object discovery problem*[12] and thus the recognition and maintenance of IDD [22]. Other related work focus on partial or global schema integration with the discovery of semantic knowledge assumed to be done in prior[21]. The FCSI is the first approach towards a federative system[19] which aims for such an intelligent support of a context-based recognition of interrelated data in autonomous databases. For this purpose the FCSI uses in particular methods from the different research areas of terminological knowledge representation and reasoning [20] as well as distributed artificial intelligence (DAI)[19].

At each agent a *descriptive, terminological meta-knowledge and domain information representation* DIT *on top of the local conceptual database schema* is built by using user-specified intentional scripts. In addition, a respectively instantiated aspect world W enables a more flexible representation of schema object aspects. According to a prior known set of common terminological primitives cpc of respective local DITs remote formal *terminological classification* then serves as a basis for the *discovery of intentional IDD*s (i-IDD) between objects of different schemas just with respect to some or all of their related intentional aspects specified by user. At the same time, projections on respective associated aspect valuations at schema and state level isps result in proposals for *possible interdatabase schema assertions* (IDSA). A set of IDSA between yet known and available intentionally related schema objects is hence infered by integrating the actual intentional state filters in the isps of detected i-IDD into a respective integrity constraint.

Context-based data sharing at schema and state level is thus determined by prior recognition of intentional relevance at terminological level and restricted on just the respective aspect valuation specified by user. As agents knowledge grows it provides the user with all its logically deducted i-IDDs and reports them as well as all respective IDSA proposals. Thus, in contrast to FDBS there is no need and prior possibility for the user to browse first through all remote available export schema structures for finding

some possibly related data. Moreover, such kind of context-based discovery of interdatabase dependencies and respective data sharing possibilities supports in particular strict association autonomy. In order to organize a *cooperative recognition process* of i-IDDs methods for flexible *utilitaristic coalition building among the autonomous agents of the FCSI* are currently investigated. Thus, no global information agent [3] or central mediator agent [4] exists. For supporting the research on the FCSI we recently started to develop an interactive development environment for the specification and simulation of agent systems IDEAS-1.
The remainder of this paper is organized as follows. In section 2 a short overview on the functionality and architecture of each FCSI agent is given. Section 3 concludes and gives an outlook on future research on the FCSI.

2 FCSI : Functionality and architecture

As reported in [18] according to a set of user-specified *intentional scripts* on some exportable own schema object structures a local *domain information terminology* DIT and its actually instantiated schema aspect world W on top of the conceptual database schema is incrementally built at each agent. For this purpose an intentional terminological formalism ITF as well as a schema aspect assertional formalism SAF as its conservative extension[20] is used. It provides most usual term-forming operators like conjunction, number restriction and value restriction as well as atomic concept negation. [1]. For reasons of space limitation for a comprehensive introduction to hybrid terminological knowledge representation and reasoning we refer the reader to [20]. The main steps for the construction of actual terminological knowledge base DIT/W are as follows.

- for each available schema object $o_k \in SObj$: **set intentional script** $intsc_k$
  (specify frame with set of pairs of aspect entries $i_x \in intset_k \subseteq Int$ using a script language ISL)
  terminologic: express some intentional semantic aspect i_x on o_k by
  terminologically defining intentional reference concepts and actual instantiation with respect to its intended linking to o_k;
  aspect valuation (intentional filter): set meta-structure $\mathsf{isp}(o_k, i_x) := < o_k, CP^{i_x}, JC^{i_x} >$ [2].
- **construct DIT/W**: for all given scripts
  - compile expression within each entry into equivalent ITF statements (for term introductions in DIT) and SAF assertions (for respectively instantiated *intentional aspect reference objects* $intob_x \in RObj$ in W); [3]
  - build directed, acyclic concept graph CT_{DIT} reflecting immediate term-subsumption relationships and realize if W is consistent with respect to the vocabulary DIT by using terminological inference algorithms: `tsub`, `classify`, `realize`[20] [4]

It is now possible for the agents to automatically detect a set of *intentional interdatabase dependencies* i-IDD between two objects o_1,o_2 in different schemas with respect to some or all of their intentional aspects. Each considered semantic aspect on o_k is terminologically represented by the abstraction of its respective reference object *intob* in W. Formal terminological comparison of the respective *intentional abstractions of schema objects* [5] then determines their kind of contextual agreement[31] [6].

$\mathbf{p-intsub(o_1, o_2, M, N)} :\Leftrightarrow$
$(\ \forall i_x \in M \subseteq intset_1 \exists i_y \in N \subseteq intset_2$:
$tsub(intabst(o_2, i_y), intabst(o_1, i_x)) \ ) \wedge$
$(\ \forall i_y \in N \ \exists i_x \in M$:
$tsub(intabst(o_2, i_y), intabst(o_1, i_x)) \ )$

$\mathbf{c-intsub(o_1, o_2, M, N)} :\Leftrightarrow$ p-intsub with M $= intset_1$
$\mathbf{inteq(o_1, o_2)} :\Leftrightarrow c-inteq(o_1, o_2, intset_1, intset_2)$
$\mathbf{intdis(o_1, o_2)} :\Leftrightarrow c-intdis(o_1, o_2, intset_1, intset_2)$
$\mathbf{intsub(o_1, o_2)} :\Leftrightarrow c-intsub(o_1, o_2, intset_1, intset_2)$

[1] A formal description of the respective functional interface between the schema and its terminological meta-knowledge base can be found in [16]. It enables in particular a flexible modeling of distinct semantic aspects on one schema object the user has in mind for external representation. Note: the expressivity of ITF/SAF restricts that of schema structure aspects the user intends to represent, besides it depends on the user how close he models the DIT/W wrt. the given schema (use of same names, pure data structure copy cf.[2]). We currently use the terminological language NTF/AF proposed in [20] as ITF/SAF. Several alternatives for this choice like e.g. ALC [30][13] or ALCF(D) [1] are currently under consideration wrt. some efficient implementation of respective algorithms for hybrid terminological reasoning within each agent. In order to achieve homogeneous IDSA we currently presume an EER data model[14] as canonical data model for each conceptual component schema[27],

Note also that linking of some schema structures into a local agent knowledge base ontology [4] DIT is flexibly done by using reference objects (of DIT ontology concepts) in W independently from used data model.

[2] CP^{i_x} denotes the set of characteristic *structure parts* and JC^{i_x} a *state constraint or function* actually specified by user with respect to an aspect i_x of o_k he has in mind. Detailed examples are in [16].

[3] The ITF/SAF is thus used as logical language with model-theoretical semantics on the domain $(SObj \times Int) \cup RObj$ of schema aspect references. This is in contrast to [6][29].

[4] Term subsumption is decidable for NTF/AF[20] as well as for ALC and its several derivatives[30]. As in most practical terminological systems for reasons of efficiency actually used `tsub` is polynomial but incomplete as this is the case also for `realize`. `tsub` is reflexive and transitive.

[5] $\mathrm{abst}(intob_x)$ of reference object *intob*: conjunctive term built by using all actual *intob* assertions in W; $intabst(o_1, i_x) := expand(abst(intob_x))$, contains only primitive components (cf.[20]).

[6] If e.g. o_1 and o_2 are differently named the recognition of $inteq(o_1, o_2)$ indicates them to be *intentional synonym with respect to their whole actual aspect sets*. Intentional homonyms can be found using intdis-relations.

The complete set of i-IDDs can be found in [16]. In accordance with [24], to enable mutual intentional understanding of received foreign terms a set $cpc_{i,j}$ of corresponding local primitive components is used between communicating agents a_i, a_j. The *recognition process of i-IDDs* executed at each agent a_i locally includes the following main steps:

- **receive** task to find some (desired) i-IDD on term $intabst(o_1, i_x)$ from agent a_j;
- **substitute**: consider received term as own concept introduction sac_i^x in the local DIT
  substitute primitive components in $intabst(o_1, i_x)$ by corresponding local ones using $cpc_{i,j}$
- **classify** sac_i^x into the local concept taxonomy CT_{DIT}:
  reduce object search space by locating some concepts c with respect to desired i-IDD;
- **find** relevant reference objects in W :
  compare the abstraction of each actual instance $intob_y$ of c with $intabst(o_1, i_x)$ wrt i-IDD;
- **project** on local schema:
  determine the particular schema aspect (o_2, i_y) the founded $intob_y$ refers to;
  check availability of i_y; get valuation $\mathsf{isp}(o_2, i_y)$; report i-IDD to a_j; (send CP^{i_y})

Considering now possible *context-based data sharing*, any request of the agent a_j on yet known intentionally related schema object o_2 to a_i is locally restricted at schema as well as filtered at state level by the respective aspect valuation $\mathsf{isp}(o_2, i_y)$ [7]. The task of indicating some *interdatabase schema assertion IDSA* is supported by relating the intentional state filters JC of an i-IDD as a respective integrity constraint using rules like

if $p - inteq(o_k, o_j, \{i_x\}, \{i_y\})$ **then** `propose` $[JC^{i_x} \Leftrightarrow JC^{i_y}]$.
if $p - intsub(o_k, o_j, \{i_x\}, \{i_y\})$ **then** `propose`$[JC^{i_x} \Rightarrow JC^{i_y}]$; [8]

However, the final decision of accepting such discovered set of IDSAs proposed by the agent remains with the humans as also stated in [28]b.
The main task of each FCSI agent is to support the user in discovering the available information space with respect to its own local domain of interest DIT. This can be followed up by the agent through detecting some i-IDDs which are requested by the user as one of its task [9]. Each agent is coarsely structured as follows. A central *nucleus* NM is responsible for internal task management, coalition negotiation and action scheduling with respect to the execution of some task received from local user or other agents. All user-specified intentional scripts on schema object aspects are maintained by the *Script Manager* SM. The terminological knowledge DIT/W on the local schema is then constructed and maintained by the *TKM module.* A *Recognition module* RM infers i-IDDs by joining the *TKM* and projecting on respective valuations isp stored at the SM. It is also responsible for maintaining these detected i-IDD in its *belief base* which includes in particular all logically deducted i-IDD facts (PROLOG facts). Any request on some particular or the existence of any i-IDD with other agents are handled by the RM (using PROLOG). IDSA proposals are reported by a RM submodule. An appropiately partitioned local blackboard is used for agents announcements and constrained local knowledge propagation between actually committed agents. FCSI *coalitions* are group of agents which are mutually committed to execute some of their tasks for a limited period of time under certain negotiated constraints. Since each agent is individual self-motivated and autonomous, it decides by itself about the kind of joining such a coalition considering its own possible Shapley-value based *utility of participation*[15][25] We currently investigate several different types of coalition utilities and thus possible types of coalition between agents with respect to our application. One such utility is based on the amount of task interactions constituted by detected i-IDDs satisfying received as well as local own FIND-task goals.
There is no need for any central federal dictionary or federated schema: Epistemic knowledge of an agent is stored in its local belief set[32] and maintained locally. Cooperation dialogue can be modeled by using a set of cooperation primitives, objects and constraints[19] in compliance with KQML[11]. Research on the FCSI mainly aims for the description of such coalition-based cooperation for i-IDD recognition.

3 Conclusion

The proposed approach of the FCSI enables the user to discover some intentional relevant data without the need to browse through all available schema structures first. Such context-based recognition

[7] Ex.: a_j: `request-for` o_2 `wrt` i_y; a_i: (EER-DML) `retrieve` x($\in CP^{i_y}$) `from` o_2 `where` JC^{i_y}

[8] $[\,] : \forall a_{type(o_k)}, b_{type(o_j)} : (a = b) \Rightarrow (JC^{i_x}\ rel\ JC^{i_y})$ closed; instance identification by corresponding keys;
If o_1's intentional aspect is related to an aspect of o_2 then assume the same relation for their user-specified justification at state level. Each set of IDSA proposals give hints for possible object domain mappings and can be extended as agent's i-IDD set (knowledge) grows. Rules for IDSA proposals are further investigated in a more detail in [16].

[9] Tasks are determined by user-specified intentional queries (use common language CIRCL) with respect to desired i-IDD and IDSA recognition (FIND) as well as context-based data sharing (REQUEST-FOR).

is exclusively done in the FCSI at information type level, i.e. by formal terminological classification of some considered aspect term independent from its actually attached and sofar protected data structures. This is similar to the idea of incrementally building and using some shared ontology for contextual interchange[31][4] respecting association autonomy[27]. Thus, possible data sharing as well as proposing IDSAs is determined by the prior success and kind of such terminological aspect classification and restricted on the respective aspect valuations specified by user. The agent provides the user with all locally deducted i-IDDs and reports respective possible IDSAs. These can be used for final decisions in specification of IDDs by human[28]b.
There has been only little related research on using formal terminological classification for the object discovery problem like in [6]. Recent works on system approaches which have influenced the FCSI approach are in particular [5],[21],[4] and [12]. Partially related works are [6],[2] and [3] [10].

Ongoing research on the FCSI includes the following main topics:
- investigations on formal FCSI agent description[32] and possible implementation with IDEAS-1
- application of utilitaristic coalition building for organizing cooperation within the FCSI[15]

Acknowledgements: I would like to thank Prof. Dr. D. Klusch, Prof. Dr. P. Kandzia, T. Prey, J. Wappler and A. Klusch for reviewing an earlier draft of this paper and giving many helpfully hints and advices.

References

[1] Baader,F., Hanschke,P., 1992,"Extensions of concept languages for a mechanical engineering application", in: LNAI 671

[2] Beck, H.W., et al., 1989,"Classification as a query processing technique in the CANDIDE SDM",IEEE Computer

[3] Barbuceanu,M., Fox,M.S., 1994,"The information agent: an infrastructure for collaboration in the integrated enterprise", Proc. CKBS-94, Keele(UK)

[4] Behrendt,W., et al., 1993,"Using an intelligent agent to mediate multibase information access", Proc. CKBS-93, Keele(UK)

[5] Bouguettaya,A., 1992,"A dynamic framework for interoperability in large MDB", Ph.D. thesis

[6] Catarci,T., Lenzerini,M.,"Interschema knowledge in cooperative IS", Proc. ICIS-93, Rotterdam

[7] Ceri, S., Widom, J., 1992,"Managing semantic heterogeneity with production rules and persistent queues", Politecnico Milano TR 92-078

[8] Duong,T. et al., 1993, "A framework of flexible and dynamic integration for MDBS", Proc. ICIS-93, Rotterdam

[9] Elmagarmid, E., Zhang, A., 1992,"Enforceable interdatabase constraints in combining multiple autonomous databases", Purdue Tech. Rep. CSD-TR-92-008

[10] Fox,J., et al., 1993,"Acceptability of arguments as logical uncertainty", LNCS 747

[11] Genesereth,M. et al.,"Specification of the KQML Agent-Communication Language", Draft Spec. Rep. 6/1993, DARPA Knowledge Sharing Initiative EIWG

[12] Hammer,J., et al., 1993,"Object discovery and unification in FDBS", in [23]b

[13] Hollunder,B., Nutt,W., 1990,"Subsumption algorithms for concept languages", DFKI-Res.Rep. RR-90-04

[14] Kandzia,P., Klein,H.-J., 1993,"Theoretische Grundlagen relationaler Datenbanken", BI

[15] Ketchpel, S., 1993,"Coalition formation among autonomous agents", Proc. MAAMAW-93

[16] Klusch, M., "Ein föderatives Zellsystem FCSI zur Unterstützung einer kontextbasierte Erkennung plausibler Interdatenbankabhängigkeiten", Ph.D. thesis, (in preparation)

[10] For reasons of space limitation for a detailed discussion of related works cf. [16].

[17] Klusch, M., 1994,"Towards a Federative System FCSI for a context-based recognition of plausible Interdatabase Dependencies", Proc. 6. GI-Workshop on 'Foundations of Databases', Bad Helmstedt

[18] Klusch, M., 1995,"Towards a Federative System FCSI for a context-based discovery of Interdatabase Dependencies", Proc. ETCE-95 Intern. Sympos. on 'Knowledge-based Systems in Engineering', Houston (USA), to appear

[19] Mueller, J. (Hrsg.), 1993, "Verteilte Künstliche Intelligenz", BI

[20] Nebel, B., 1990, "Reasoning and revision in hybrid representation systems", LNAI 422, Springer

[21] Ouksel,A., Naiman,C.F., 1993,"Coordinating context building in heterogeneous IS",IJICIS

[22] Papazoglou,M.P. et al.,1992,"An organizational framework for intelligent cooperative IS", IJICIS 1(1)

[23] a) Proc. IEEE RIDE-93, Vienna, 1993 b) Proc. SI/DBTA-Wshp., Fribourg, 1993

[24] Sabah, G., 1993, "Knowledge Representation and Natural Language Understanding", AICOM 6(3/4)

[25] Shechory, O., Kraus, S., 1994, "Coalition formation among autonomous agents (preliminary report)"

[26] Sheth,A., Kashyap,V., 1992,"So far schematically yet so near semantically", Proc. IFIP TC2/WG2.6

[27] Sheth,A., Larson,J.A., 1990,"Federated database systems for managing distributed, heterogeneous and autonomous DBS", ACM CS 22(3)

[28] Sheth, A., et al., 1991,"Specifying interdatabase dependencies in a MDB environment", IEEE Computer (see also b) Bellcore TM-STS-018609/1)

[29] Sheth, A., et al., 1991,"On applying classification to schema integration", Proc. IEEE 1.Wshp Interop. MDBS, Kyoto(Japan)

[30] Smolka,G., Schmidt-Schauß,M., 1991,"Attributive concept description woith complements", AI 48

[31] Siegel, M., Madnick, S.E., 1991, "Contextual interchange: sharing the meaning of data", ACM SIGMOD 20(4)

[32] Woolridge, M., 1993, "MYWORLD: an agent-oriented testbed for DAI", Proc. CKBS-93, Keele(UK)