Social Contexts and the Probabilistic Fusion and Ranking of Opinions: Towards a Social Semantics for the Semantic Web*

Matthias Nickles[1] and Ruth Cobos[2]

[1] AI/Cognition Group, Computer Science Department, Technical University of Munich, D-85748 Garching b. München, Germany, nickles@cs.tum.edu

[2] Departamento de Ingeniería Informática, Universidad Autónoma de Madrid, 28049 - Madrid, Spain, ruth.cobos@uam.es

Abstract. In the (Semantic) Web, the existence or producibility of certain, consensually agreed or authoritative knowledge cannot be assumed, and criteria to judge the trustability and reputation of knowledge sources may not be given. These issues give rise to the formalization of web information in terms of heterogeneous and possibly inconsistent public assertions and intentions, providing valuable meta-information in contemporary application fields, like open or distributed ontologies, social software, ranking and recommender systems, and domains with a high amount of controversies, such as politics and culture. As an approach towards this, we introduce a lean, intuitive formalism for the Semantic Web which allows for the explicit representation of semantic heterogeneity by means of so-called *social contexts*, and optionally for the probabilistic aggregation and social rating of possibly uncertain or contradictious assertions. Inter alia, this allows to stochastically generalize multiple assertions (yielding complexity reduction), and generalizes the concept of *folksonomies* to any ontologies and description logic knowledge bases emergent from social choice processes.
Keywords: Semantic Web, OWL, Uncertainty, Information Integration, Context Logic, Social Choice

1 Introduction

Information found in open environments like the web can usually not be treated as objective, certain knowledge directly, and also not as sincere beliefs (due to the mental opaqueness of the autonomous information sources). But only very few approaches to the semantic modeling of what could be called "opinions" or "public assertions", which are neither (real) beliefs nor objective knowledge, have emerged, mostly in the field of distributed artificial intelligence [10, 11]. Instead, most formal approaches to knowledge representation and reasoning handle logical inconsistencies and information source controversies as something which should be avoided or filtered out using criteria such as trust and provenance. Against that, we argue that making (meta-)knowledge about the social, heterogeneous and controversial nature of web information explicit can be extremely useful [7] - e.g., to gain a picture of the opinion landscape in controversial domains such as politics, for subsequent decision making and conflict resolution, for the acquisition and ranking of information from multiple, possibly dissent sources, and not at last for tasks like the learning whom (not) to trust. Such (meta-)knowledge is especially crucial in domains with a strong viewpoint competition and difficult or impossible consensus finding like politics, product assessment and culture, and in current and forthcoming Semantic Web applications which support explicitly or implicitly people interaction, like (semantic) blogging, discussion forums, collaborative tagging and folksonomies, and social computing

* An extended variant of this paper is published as Technical Report FKI-254-06, AI/Cognition Group, Technical University of Munich, 2006.

in general. Approaching this issue, this work presents a lean approach to the formal representation of semantical heterogeneity by means of *social contexts* and the *social rating* of possibly contradictious or uncertain assertions via opinion weighting and probabilistic opinion aggregation.

2 Modeling heterogeneous viewpoints using social contexts

2.1 Modeling social structures

Our technical approach is based on providing an interrelationship of a social ontology or social knowledge base (KB) for the description of social concepts and individuals (like persons, agents and organizations, and maybe their relationships) on the one hand, and a set of possibly controversial or uncertain statements (axioms and facts) on the other. Special terms consisting of names from the social ontology/KB then identify so-called *social contexts* for the contextualization and optionally the fusion of semantically heterogeneous statements. This amounts to a technique which makes use of the context-driven partitioning of the respective web language semantics analogously to the approach presented in, e.g., [2, 4].
There is no canonical social ontology to be used with our approach. Basically any ontology could be used as long as it provides concepts and instances for communication participants like "Author", "Publisher" or "Reader", or, most simply, "Actor". Of course, the following approach would also work in case the participants are stated indirectly in form of the web locations or resources they use to articulate themselves. But our approach suggests that information sources shall even then be seen as autonomous, communicating actors. The following example ontology fragment will do in this work:

Definition 1: Social ontology SO

$Source \sqsubseteq Actor$, $Addressee \sqsubseteq Actor$ $Source(tina)$, $Source(tom)$, $Source(tim)$,
$Addressee(tina)$, $Addressee(tom)$, $Addressee(tim)$,
$CA(assertion)$, $CA(information)$, $CA(publicIntention)$,
$CA(fusedInformation)$, $CA(fusedAssertion)$

At this, *Source* and *Addressee* are the classes of the participating actors, whereby *Source* can be any kind of information source, like a person, an organization, a document, or a web service, or the holder of an intention. *assertion* denotes an *ostensibly* positive (i.e., approving) *communication attitude* (CA) regarding a certain statement, and at the same time the *ostensible* intention to make the addressee(-s) adopt a positive attitude also (e.g., "This product is the best-buy!"). This corresponds more or less to the communication act semantics we have introduced in detail in [10, 11], and to Grice's conceptualization of speech acts as communications of intentions. *information* is the same as *assertion*, but without the intention to make others approve the resp. statement too (e.g., "Personally, I believe in god, but I respect your agnosticism"). *information*s and *assertion*s are also called "opinions" in this work. *publicIntention* finally is the communication attitude of ostensibly desiring that a statement shall become true. The attitude of *requesting* something from another actor would be a subtype of *publicIntention*. Likewise, the attitude of *denying* something can simply be substituted by the positive attitude towards the negation of the denied statement. The three attitudes defined in SO should be sufficient to represent most information, publishing and desiring acts on the internet.
Note that *assertion*, *information* and *publicIntention* are no propositional attitudes in the usual mentalistic sense, as they do not need to correspond to any sincere beliefs or intentions of the actors. Instead, they are possibly insincere *communication* or *social* attitudes exhibited

in a synchronous or asynchronous communication process like that of the publishing of information on the web. As a consequence, they can not be treated like their mental counterparts. E.g., an actor might hold the opinion ϕ towards addressee one and at the same time $\neg\phi$ informing addressee two (while believing neither ϕ nor $\neg\phi$ privately). As another example, opinions can even be bought, in contrast to sincere beliefs: It is known that opinions uttered in, e.g., web blogs have sometimes been payed for by advertising agencies. In some sense, even *all* information on the web is "just" opinion due to the absence of a commonly accepted truth authority. *fusedInformation* and *fusedAssertion* will be described below.

2.2 Social contexts and social semantics

Contexts (aka microtheories) have been widely used in AI since the early nineties, originally intended by McCarthy as a replacement of modal logic. [1, 2] propose a context operator $ist(context, statement)$ which denotes that *statement* is true ("ist") within *context*. Building upon [2, 4], we will use the notation of context to express that certain statements are being publicly asserted (informed about, ostensibly intended to become true, denied...) on the web by some information-*Source*(s), optionally facing some specific *Addressee*(s) (the latter implies that our use of the term "public" optionally comprises "limited publics" in form of closed social groups also). Thus, such *social contexts* model the *social semantics* of the contextualized information. Here, the term "social semantics" has a twofold meaning itself: First, it refers to the *communicative function* of information published on the web (essentially, our contexts correspond to kinds of speech acts). Second, a social context can optionally denote the meaning of a certain statement ascribed by *multiple* actors using some aggregation rule, e.g., the degree of truth assigned via consensus finding or voting, or other kinds of social choice.

Definition 2: Social contexts

A *social context* is formally defined as a pair (id, c) with $id \in Id$,
$Id = (Source^{\mathrm{N}} \times ... \times Source^{\mathrm{N}}, Addressee^{\mathrm{N}} \times ... \times Adressee^{\mathrm{N}}, CA^{\mathrm{N}})$, c being a set of mutually consistent OWL expressions (precisely: axioms and facts as defined in the next section), and *Source*, *Adressee* and *CA* being the respective categories in *SO*. The X^{N} denote sets of names of the individuals within the respective categories (so we assume an extensional denotation of social concepts like persons and groups, as far as the contained individuals are required for context specification), e.g., $Source^{\mathrm{N}} = \{"tina", "tom", "tim"\}$. *id* is called the *context identifier* (sometimes called "context" also for short). We use the following syntax for context identifiers:

$$^{attitude}_{source_1,...,source_n \rightarrowtail addresse_1,...,addressee_n}$$

As an abbreviation, we define $^{attitude}_{source_1,...,source_n} = ^{attitude}_{source_1,...,source_n \rightarrowtail Addresse^{\mathrm{N}}}$, i.e., the attitude is here addressed to the group of *all* possible addressees like it is the case with information found on an ordinary web page. But note that at the same time a certain source can hold mutually inconsistent attitudes even towards different members or subgroups of $Addresse^{\mathrm{N}}$ (but not towards the same addressee).

3 A description logic with social contexts

We settle on the $\mathcal{SHOIN}(D)$ description logic (over data types D), because ontology entailment in the current quasi-standard OWL DL can be reduced to $\mathcal{SHOIN}(D)$ KB satisfiability [14]. Since we don't make use of any special features of this specific description language, our approach could trivially be adapted to any other description language or OWL variant, or even first-order logic and RDF(S).

Definition 3: $\mathcal{SHOIN}(D)$-ontologies

The context-free grammar of $\mathcal{SHOIN}(D)$ is defined as follows:

$$C \rightarrow A|\neg C|C_1 \sqcap C_2||C_1 \sqcup C_2|\exists R.C|\forall r.C$$
$$\quad | \geq nS| \leq nS|\{a_1, ..., a_n\}| \geq nT| \leq nT|\exists T_1, ..., T_n.D|\forall T_1, ..., T_n.D$$
$$D \rightarrow d|\{c_1, ..., c_n\}.$$

At this, C denote *concepts*, A denote *atomic concepts*, R denote *abstract roles* (relationships), S denote *abstract simple roles*, the T_i denote *concrete roles*, d denotes a concrete *domain predicate*, and the a_i, c_i denote abstract / concrete *individuals*.
A $\mathcal{SHOIN}(D)-$*ontology* is a finite set of TBox and ABox axioms/facts $C_1 \sqsubseteq C_2$ (inclusion of concepts), $Trans(R)$ (transitivity), $R \sqsubseteq S$, $T \sqsubseteq U$ (role inclusion), $C(a)$ (concept assertion), $R(a, b)$ (role assertion), $a = b$ (equality), and $a \neq b$ (inequality). For lack of space, please find the semantics of $\mathcal{SHOIN}(D)$ in [14].

Definition 4: SOC-OWL

Introducing ontologies and at the same time description logic knowledge bases with social contexts, we define *SOC-OWL* (*Social-Context-OWL*) similarly to C-OWL [4]. In Section 4, an advanced language P-SOC-OWL will be introduced.
A SOC-OWL ontology/KB is a finite set $\{(id, s) : id \in Id, s \in AF\} \cup AF^i \cup B$, with AF being the set of all $\mathcal{SHOIN}(D)$ axioms/facts, AF^i being such axioms/facts but with concepts, individuals and roles directly indexed with social contexts (i.e., $AF^i = \{(id_i, C_h) \sqsubseteq (id_j, C_k), (id_i, a_h) = (id_j, a_k), ... : id_i, id_j \in Id\}$), and B being a set of *bridge rules* (see 3.1). Id is the set of all social context identifiers according to the social ontology SO (cf. Definition 1). The s within (id, s) (i.e., plain OWL-DL axioms/facts) are called *inner statements* which are said to "be true (or intended in case of *publicIntention*) within the respective context". Example (with multiple axioms/facts per row and (id, a) written as $id\ a$):

$ControversialPerson(columbus)$ $\quad$ ${}^{assertion}_{tina \rightarrowtail tim,tom}Hero(columbus)$
${}^{assertion}_{tim,tom \rightarrowtail tina}(\neg Hero)(columbus)$ $\quad$ ${}^{assertion}_{tim,tom \rightarrowtail tina}Exploiter(columbus)$

This SOC-OWL ontology (modeling as a whole somewhat a neutral point of view, like taken by an ideal Wikipedia article) expresses that the (fictive) persons Tim and Tom hold the opinion towards Tina that Christopher Columbus was not a hero but an exploiter (of the natives), while Tina does allegedly believe that the opposite is true. But there is consensus of the whole group that Christopher Columbus is a controversial person. Notice that without explicit further constraints such as the specified in 3.1 and 3.2, different social contexts are logically fully separated. E.g., from the above ontology it could *not* be deduced that ${}^{information}_{tina \rightarrowtail tim}ControversialPerson(columbus)$, because $ControversialPerson(columbus)$ as an abbreviation of ${}^{information}_{tina,tim,tom \rightarrowtail tina,tim,tom}ControversialPerson(columbus)$ in the example above is uttered/addressed *exactly* by/to *the single social group of all* participants. This principle allows to model the realistic case that someone conforms with some group opinion, but states some inconsistent opinion towards other groups (even a subgroup of the former group). Of course the co-presence of two or more social contexts which indicate that a certain actor is insincere (as it would be the case if ${}^{assertion}_{tina \rightarrowtail tim}(\neg C)(x)$ and ${}^{assertion}_{tina \rightarrowtail tom}C(x)$ were contained within the same SOC-OWL ontology, which would be perfectly legal) could usually not be acquired directly from the web, since such actors would likely exhibit inconsistent opinions using different nicknames. Instead, some social reasoning or social data mining techniques would be required to obtain such SOC-OWL knowledge.
Obviously, each contextualized SOC-OWL statement $(contextId, statement)$ corresponds to

the "classic" [1, 2] context statement $ist(context, statement)$. But unfortunately, such an *ist* operator cannot simply be made a first-class object of our language (which would allow for the nesting of context expressions), at least not without getting into trouble defining a semantics of the language, or without making the semantics as shallow as that of RDF-style reification. Instead, we allow for *bridge rules* and meta-axioms in order to interrelate social contexts.
The core idea underlying the following semantics of SOC-OWL is to group the axioms/facts according their social contexts, and to give each context its own interpretation function and domain within the model-based semantics, corresponding to the approach presented in [4]. In addition, we will provide meta-axioms (constraints) and bridge rules in order to state the relationships among the various communication attitudes (somewhat similarly to modal logic axiom schemes such as the well-known KD45 axioms of modal belief logic), and to allow for the interrelation of different attitudes, even across different contexts. E.g., we would like to express that a communication attitude such as ${}^{assertion}_{tina \rightarrowtail tim,tom}(\neg Exploiter)(columbus)$ implies (intuitively) ${}^{publicIntention}_{tina}({}^{information}_{tim,tom \rightarrowtail tina}(\neg Exploiter)(columbus))$, i.e., that Tina not only expresses her ostensible beliefs, but also ostensibly intends that others adopt her opinion.

Definition 5: Interpretation of SOC-OWL

A SOC-OWL *interpretation* is a pair $(I, \{e_{i,j}\}_{i,j \in Id})$ with $I = \{I_{id}\}$ being a set of *local interpretations* I_{id}, with each $I_{id} = \langle \triangle^{I_{id}}, (.)^{I_{id}} \rangle$, $id \in Id$. $e_{i,j} \subseteq \triangle^{I_i} \times \triangle^{I_j}$ is a relation of two *local domains* $\triangle^{I_{id}}$ ($e_{i,j}$ is required for the definition of bridge rules in B (Definition 4) as explained later in 3.1). $(.)^{I_{id}}$ maps individuals, concepts and roles to elements (resp. subsets or the products thereof) of the domain $\triangle^{I_{id}}$.
To make use of this interpretation, contextualized statements of SOC-OWL impose a grouping of the concepts, roles and individuals within the inner statements into sets C_{id}, R_{id} and c_{id} [4]. This is done in order to "localize" the names of concepts, individuals and roles, i.e., to attach to them the resp. local interpretation function I_id corresponding to the resp. social context denoted by $id \in Id$:
Concretely, the sets C_{id}, R_{id} and c_{id} are defined inductively by assigning the concepts, individuals and role names appearing within the *statement* part of each SOC-OWL axiom/fact $(context_{Id}, statement)$ to the respective set C_{id}, c_{id} or R_{id}. With this, the interpretation of concepts, individuals etc. is as follows:

$C^{I_{id}} =$ any subset of $\triangle^{I_{id}}$ for $C \in C_{id}$
$(C_1 \sqcap C_2)^{I_{id}} = C_1^{I_{id}} \cap C_2^{I_{id}}$ for $C_1, C_2 \in C_{id}$
$(C_1 \sqcup C_2)^{I_{id}} = C_1^{I_{id}} \cup C_2^{I_{id}}$ for $C_1, C_2 \in C_{id}$
$(\neg C)^{I_{id}} = \triangle^{I_{id}} \setminus C^{I_{id}}$ for $C \in C_{id}$
$(\exists R.C)^{I_{id}} = \{x \in \triangle^{I_{id}} : \exists y : (x, y) \in R^{I_{id}} \wedge y \in C^{I_{id}}$ for $C \in C_{id}, R \in R_{id}$
$(\forall R.C)^{I_{id}} = \{x \in \triangle^{I_{id}} : \forall y : (x, y) \in R^{I_{id}} \rightarrow y \in C^{I_{id}}$ for $C \in C_{id}, R \in R_{id}$
$c^{I_{id}} =$ any element of $\triangle^{I_{id}}$, for $c \in c_{id}$
(Interpretation of concrete roles T analogously)

Satisfiability and decidability

Given a SOC-OWL interpretation I, I is said to *satisfy* a (contextualized) statement ϕ ($I \models \phi$) if there exists an $id \in Id$ such that $I_{id} \models \phi$. A SOC-OWL ontology/KB (or statement set) is then said to be satisfied if I satisfies each statement within the ontology/KB (or statement set). $I_{id} \models (id, C_1 \sqsubseteq C_2)$ iff $C_1^{I_{id}} \subseteq C_2^{I_{id}}$ etc., as with $\mathcal{SHOIN}(D)$ (but indexed). Note that using this extension, the inherited semantics and decidability of $\mathcal{SHOIN}(D)$ and C-OWL remain unaffected in SOC-OWL within each context, since the new interpretation function simply decomposes the domain and the set of concepts etc. into local "interpretation modules" corresponding to the contexts.

3.1 Bridge rules and cross-context mappings

According to Definition 4, a SOC-OWL ontology can optionally comprise bridge rules [4] B and various stronger relationships among classes, individuals and roles from different contexts. As an example, consider

$(context_i, x) \xrightarrow{\equiv} (context_j, y)$ in B, with x, y being concepts, individuals or roles.

Informally, such a bridge rule states that the x and y denote corresponding elements even though they belong to different contexts $context_i, context_j$.

With, e.g., $({}^{assertion}_{tina}, columbus) \xrightarrow{\equiv} ({}^{assertion}_{tim,tom}, columbus)$ the interpretations of the "two Columbuses" would abstractly refer to the same object. Analogously, $\xrightarrow{\sqsubseteq}$ and $\xrightarrow{\perp}$ state that the first concept is more specific than the second resp. that both concepts are disjoint. These relationships are given by the relation $e_{i,j}$ (Definition 5).

Formally: $I \models (context_i, x) \xrightarrow{\equiv} (context_j, y)$ iff $e_{i,j}(x^{I_i}) = y^{I_j})$ (resp. $e_{i,j}(x^{I_i}) \subseteq y^{I_j}$ and $e_{i,j}(x^{I_i}) \cap y^{I_j} = \emptyset$).

For lack of space, please find details and analogously defined further bridge rules in [4].

A much stronger kind of relationship is stated by the syntax constructs where a concept, individual or role is directly indexed with a social context, as, e.g., in $(context_i, x) = (context_j, y)$, with x, y being concepts, individuals or roles.

Formally: $I \models (context_i, x) = (context_j, y)$ iff $x^{I_i} = y^{I_j}$ (analogously for $\sqsubseteq, \perp$ etc).

3.2 Meta-axioms

We state now some constraints regarding the social meaning of contexts, which will later be extended with (PMA5).

(MA1) Actively asserting an opinion implies in our framework the intention of the source that the addressee(-s) adopt the asserted statement. With nested social contexts, we could formalize this using ${}^{assertion}_{s_1,...,s_n \rightarrowtail a_1,...,a_m}\varphi \rightarrow ({}^{publicIntention}_{s_1,...,s_n \rightarrowtail a_1,...,a_m}({}^{information}_{a_1,...,a_m \rightarrowtail s_1,...,s_n}\varphi)$. To avoid such "strong" nesting in order to lower decidability complexity, we propose the following as a significantly weaker replacement, which at least allows to keep track of the convictions the information sources desire in a separate context:

${}^{assertion}_{s_1,...,s_n \rightarrowtail a_1,...,a_m}\varphi$

$\rightarrow (({}^{publicIntention}_{s_1,...,s_n \rightarrowtail a_1,...,a_m} \sqcup^{assertion}_{a_1,...,a_m \rightarrowtail s_1,...,s_n}, e) = ({}^{assertion}_{s_1,...,s_n \rightarrowtail a_1,...,a_m}, e))$ for each concept, individual and role e occurring in φ.

Here, the first context is a concatenation of an intention and an assertion. The intuitive meaning is that group $s_1, ..., s_n$ intends that group $a_1, ..., a_m$ shall describe relevant individuals, categories etc. within their social context in the same way as $s_1, ..., s_n$ does wrt. "their" e's.

In terms of SOC-OWL interpretation this is:

$(I_{{}^{assertion}_{s_1,...,s_n \rightarrowtail a_1,...,a_m}} \models \varphi) \rightarrow I \models (({}^{publicIntention}_{s_1,...,s_n \rightarrowtail a_1,...,a_m} \sqcup^{assertion}_{a_1,...,a_m \rightarrowtail s_1,...,s_n}, e) = ({}^{assertion}_{s_1,...,s_n \rightarrowtail a_1,...,a_m}, e))$, for each concept, individual and role e in φ.

The next meta-axiom simply demands that assertions include the attitude of informing the addressee:

(MA2) ${}^{assertion}_{s_1,...,s_n \rightarrowtail a_1,...,a_m}\varphi \rightarrow {}^{information}_{s_1,...,s_n \rightarrowtail a_1,...,a_m}\varphi$

In this work, we can not provide a meta-theory corresponding to the KD(45) axioms of modal Belief-Desire-Intention logics (but see [10]). Instead, we only demand that the inner statements of each context are mutually consistent:

(MA3) Each set a of statements such that for a specific $context$ all $(context, a_i), a_i \in a$ are axioms/facts of the same SOC-OWL ontology, has an interpretation.

Furthermore, we demand - in accordance with many BDI-style logics - that the information/assertion contexts of a certain actor on the one hand and his intention context on the other do not overlap addressing the same set of addressees, i.e., an actor does not (ostensibly) intent what he (ostensibly) believes to be the case already:

(MA4) For each a such that $\binom{publicIntention}{source \rightarrowtail addressees}, a)$ is part of an SOC-OWL ontology o, no axiom/fact $\binom{information}{source \rightarrowtail addressees}, b)$, $b \vdash a$, is part of o (analogously for assertions).

The following constraints are *not* demanded, but could be helpful in application domains were mutual opinion consistency of subgroups is desired (we use $\bigwedge$ to abbreviate a set of SOC-OWL statements).
(MAx1) $(^{attitude}_{s_1,\ldots,s_n \rightarrowtail addressees}\ \varphi) \leftrightarrow \bigwedge_{s \in 2^{\{s_1,\cdots,s_n\}}} {}^{attitude}_{s \rightarrowtail addressees}\ \varphi$
(MAx2) $(^{attitude}_{sources \rightarrowtail a_1,\ldots,a_n} \varphi) \leftrightarrow \bigwedge_{a \in 2^{\{a_1,\cdots,a_n\}}} {}^{attitude}_{sources \rightarrowtail a} \varphi$

But we can safely aggregate seemingly consented information in a separated fusion context:
(MA5) $\bigwedge_{s \in \{s_1,\ldots,s_n\}} (I_{\substack{information \\ s \rightarrowtail addressees}} \models \varphi) \rightarrow (I_{\substack{fusedInformation \\ s_1,\ldots,s_n \rightarrowtail addressees}} \models \varphi)$ (analogously for assertions).
In general, such group opinions induce a ranking of multiple statements with the resp. rank corresponding to the size of the biggest group which supports the resp. statement (this can be used, e.g., for a majority voting on mutually inconsistent statements).

4 Social rating and social aggregation of subjective assertions

Building upon social contexts, the following extension of the previously presented logical framework is optional. It makes use of uncertainty reasoning and techniques from judgement aggregation. They allow for i) the representation of gradual strengths of *uncertain opinions* held by individuals (corresponding to subjective probabilities) and social groups, and ii) the probabilistic *fusion* of semantically heterogeneous opinions held by different actors (basically by means of voting). Since the emergence of *folksonomies* can be seen as a social categorization process (social choice in form of collaborative tagging), ii) amounts to a generalization of folksonomies to social choice of ontology and knowledge base entries in general.
This feature is also useful in case traditional techniques to ontology integration fail, e.g., if the resulting merged ontology shall be accepted by all sources, but a consensus about the merging with traditional techniques to ontology mapping and alignment could not be found, or if the complexity of a high amount of heterogeneous information needs to be reduced by means of stochastic generalization. Probabilistic fusion is furthermore helpful in case statements shall be socially ranked, i.e., put in an order according to the amount of their respective social acceptance. In contrast to heuristical or surfer-centric ways of information ranking or "knowledge ranking" such as those accomplished by most web search engines, the following approach is based on semantic opinion pooling [13].
In [9], the probabilistic extension $P-\mathcal{SHOQ}(D)$ of the $\mathcal{SHOQ}(D)$ description logic has been introduced. $\mathcal{SHOQ}(D)$ is very similar to $\mathcal{SHOIN}(D)$ and thus OWL DL, but does not allow for inverse roles. [9] shows that reasoning with $P-\mathcal{SHOQ}(D)$ is - maybe surprisingly - decidable. Instead of $P-\mathcal{SHOQ}(D)$, other Bayesian/probabilistic approaches to the Semantic Web can likely also be used as a basis for our approach, e.g., [6]. $P-\mathcal{SHOQ}(D)$ is now used to define a probabilistic variant of SOC-OWL.

Definition 6: P-SOC-OWL

A *P-SOC-OWL* ontology is defined to be a finite subset of $\{([p_l, p_u], id, a_i)\} \cup \{(id, a_i)\} \cup \{a_i\} \cup AF^i \cup B$, with $p_l, p_u \in [0,1], id \in Id, a_i \in AF$, AF being the set of all well-formed $\mathcal{SHOQ}(D)$ ontology axioms/facts, and B and AF^i as in the previous section.

The $[p_l, p_u]$ are probability intervals. Non-interval probabilities p are syntactical abbreviations of $[p, p]$. If a probability is omitted, 1 is assumed. [9] further provides some consistency axioms for probabilistic reasoning with the underlying $P-\mathcal{SHOQ}(D)$, which had to be omitted here.

Definition 7: Semantics of P-SOC-OWL

The semantics of a P-SOC-OWL ontology is simply given as a family of $P - \mathcal{SHOQ}(D)$ interpretations, each interpretation corresponding to a social context (please refer to [9] for the definition of $P - \mathcal{SHOQ}(D)$ interpretations). Formally, a P-SOC-OWL interpretation is a pair $(PI, \{e_{i,j}\}_{i,j \in Id})$ with $PI = \{(PI_{id}, \mu_{id}) : id \in Id\}$ being a set of *local probabilistic interpretations* (each denoted as Pr_{id}), each corresponding to a probabilistic interpretation of $P - \mathcal{SHOQ}(D)$ and a social context with identifier *id*. $\mu_{id} : \Delta^{I_{id}} \rightarrow [0, 1]$ is a subjective probability function, and the $\Delta^{I_{id}}$ are the domains. The relation $e_{i,j}$ (required to state bridge rules) is defined analogously to SOC-OWL. At least if reasoning is done only within each context (using the resp. interpretation), P-SOC-OWL remains decidable. Example:

$[0.5, 0.8]$: ${}^{assertion}_{tim,tom \rightarrowtail tina}Exploiter(columbus)$ 0.7: ${}^{assertion}_{tina}Hero(columbus)$
0.9: ${}^{assertion}_{tim}Hero(columbus)$

This P-SOC-OWL ontology expresses inter alia that Tim and Tom (as a group, but not necessarily individually) hold the opinion that with a probability in $[0.5, 0.8]$, Columbus was an exploiter, while Tina does (publicly) believe he was a hero with strength 0.7, and Tim believes so with strength 0.9 (i.e., his private opinion disagrees with the public group opinion of him and Tom). In order to allow for a consistent fusion of opinions, we demand the following fusion meta-axiom, which effectively states how the probabilities of *social fusion contexts* are calculated. A social fusion context is a social context with more than one opinion source and a probability which pools the probabilities which subsets of the group assign to the respective statement. This allows to specify group opinions even if group members or subgroups do knowingly not agree wrt. the resp. assertion. We propose two versions of a resp. interpretation rule:

(PMA5') $(\bigwedge_{s_i \in \{s_1,...,s_n\}} (Pr_{\substack{information \\ s_i \rightarrowtail addressees}} \models \varphi[p_i, p_i])) \rightarrow (Pr_{\substack{information \\ s_1,...,s_n \rightarrowtail addressees}} \models \varphi[p, p])$
with $p = pool^{poolingType}((p_1, ..., p_n), priorKnowledge)$. At this, $Pr_{id} \models \varphi[l, u]$ attests φ a probability within $[l, u]$ in context *id*. (Analogously for the attitude *assertion*.)

A problem with (PMA5') is that it can lead to unsatisfiability in case the derived probability p is different than a probability assigned explicitly in the ontology by the resp. group (remember that a group of agents is free to assign any truth value or probability to any proposition, using any social choice procedure). A simple work around is to use a new context *fusedInformation* (resp. *fusedAssertion*) (PMA5). Another possibility would be to introduce some kind of priority reasoning which gives priority to explicitly assigned probabilities.

(PMA5) $(\bigwedge_{s_i \in \{s_1,...,s_n\}} (Pr_{\substack{information \\ s_i \rightarrowtail addressees}} \models \varphi[p_i, p_i])) \rightarrow (Pr_{\substack{fusedInformation \\ s_1,...,s_n \rightarrowtail addressees}} \models \varphi[p, p])$ (rest as PMA5').

As for $pool^{poolingType}$, there are several possibilities: In the most simple case of "democratic" Bayesian aggregation given the absence of any opinion leader or "supra-Bayesian" [13], we define $pool^{avg}((p_1, ..., p_n), \varnothing) = \frac{\sum p_i}{n}$, i.e., $pool^{avg}$ averages over heterogeneous opinions. Using this aggregation operator, we could deduce the following: 0.8: ${}^{assertion}_{tina,tim}Hero(columbus)$. Social aggregation operators are traditionally studied in the field of *Bayesian belief aggregation* (e.g., [13, 3]) and *judgement aggregation*. The most common fusion operator extends $pool^{avg}$ with expert weights (e.g., trustability or social power degrees of the information sources):

$pool^{LinOP}((p_1, ..., p_n), (weight_1, ..., weight_n)) = \sum weight_i p_i$, with $\sum_{weight_i} = 1$. Also quite often, a geometric mean is used:
$pool^{LogOP}((p_1, ..., p_n), (weight_1, ..., weight_n)) = \kappa \prod_{i=1}^{n} p_i^{weight_i}$ (κ for normalization).
All these operators have known shortcoming (see [13] for a discussion), most prominently the so-called *impossibility results* which state that no pooling method can ever satisfy all desired properties such as systematicy.
It is also noteworthy that the operators given above do not deal with the problem of *ignorance* directly (i.e., taking account the evidence the resp. information sources have obtained, as researched in Dempster-Shafer theory). But such ignorance could be modeled using the $weight_i$ of $pool^{LinOP}$ and $pool^{LogOP}$, and possibly using probability intervals instead of single probabilities. In case opinions with probability intervals $[p_i^l, p_i^u]$ shall be fused, the described fusion operators need to be accordingly applied to the interval boundaries.
One application of such rating in form of aggregated or individual probabilities is to take the probabilities (resp. the mean values of the bounds for each interval) in order to impose an order of the axioms/facts of an ontology, so that inner statements can be directly ranked regard their degree of social acceptance, as in

0.8: ${}^{information}_{voters} innerStatement_1$ (highest social rating)
$[0.5, 0.8]$: ${}^{information}_{voters} innerStatement_2$
0.2: ${}^{information}_{voters} innerStatement_3$ (lowest social rating)

Again, such a ranking can also be easily used to transform inconsistent ordinary ontologies into consistent ontologies by a voting on the statements of the inconsistent ontology: In case there are inner statements which are mutually inconsistent, a ranking can be used to obtain a consistent ordinary (i.e., OWL DL) ontology by removing from each smallest inconsistent subset of inner statements the statements with the lowest rating until all remaining elements of each subset are mutually consistent.

5 Related works and conclusion

The goal of this work is to provide a formal social semantics of possibly contradictory assertions on the web, i.e., to state their amount of social support, their communicative emergence and dissemination, and the consensus or dissent they give rise to. Doing so, we settle on the "opinion level" where neither beliefs are visible (due to the mental opaqueness of the information sources) nor criteria for the selection of useful knowledge or semantic mappings from/among heterogenous information exist initially. This is in strong contrast to the traditional aim of information integration and evolution for the determination of some consistent, reliable "truth" obtained from the contributions of multiple sources as in *multiagent belief representation and revision* (e.g. [18]) and approaches to ontology alignment, merging and mapping. Apart from the research field of knowledge and belief integration, the storage of heterogeneous information from multiple sources also has some tradition in the fields of *data warehousing* and *federated databases*, and view-generation for distributed and enterprise database systems [8], whereby such approaches do not take a social or communication-oriented perspective. *Opinions* are treated in the area of the (non-semantic) web (e.g., *opinion mining* in natural language documents) and in (informal) knowledge management (e.g., *KnowCat* [12]). The assignment of provenance information is mostly based on *tagging* and *punning* techniques, or makes use of the very problematic reification facility found in RDF(S). Advanced approaches to provenance (e.g., [17]) are already very useful if it is required to specify who contributed some information artifact (which is also done with a similar intent on the basis of social networks), but they do not provide a logic model of the meaning of being an opinion source and other communication aspects. Precisely, they allow to specify that someone "asserts" some information, but they do not tell what asserting (requesting, denying...) actually means. [5] provides an approach to the grouping of RDF statements using contexts, but without taking into account the social (i.e., communicative) aspect, and only informally. A mature approach, focusing on the aggregation of RDF(S) graphs using contexts, was presented in [2], and [4] provides a general formal account of contexts for OWL ontologies. Independently from web-related approaches, contexts have been widely used for the modeling of distributed knowledge and so-called federated databases, see e.g. [15, 16]. To further explore and work out the new "social" perspective on uncertain information on the web modeled using contexts certainly constitutes a long-term scientific and practical endeavor of considerable complexity, with this work hopefully being a useful starting point.

Acknowledgements: This work was funded by the German Research Foundation DFG (Br609/13-1, research project "Open Ontologies and Open Knowledge Bases") and by the Spanish National Plan of R+D, project no. TSI2005-08225-C07-06.

References

1. J. L. McCarthy. Notes on formalizing context. In IJCAI, pages 555-562, 1993.
2. R. V. Guha, R. McCool, R. Fikes. Contexts for the Semantic Web. Procs. of the Third International Semantic Web Conference (ISWC-04), 2004.
3. M. Richardson, P. Domingos. Building Large Knowledge Bases by Mass Collaboration. Technical Report UW-TR-03-02-04, Dept. of CSE, University of Washington, 2003.
4. P. Bouquet, F. Giunchiglia, F. van Harmelen, L. Serafini, and H. Stuckenschmidt. C-OWL: Contextualizing Ontologies, Second International Semantic Web Conference (ISWC-03), LNCS vol. 2870, Springer Verlag, 2003.
5. G. Klyne. Contexts for RDF Information Modelling. http://www.ninebynine.org/RDFNotes/RDFContexts.html, 2000.
6. P. Costa, K. B. Laskey, K. J. Laskey. PR-OWL: A Bayesian Framework for the Semantic Web. In Procs. First Workshop on Uncertainty Reasoning for the Semantic Web (URSW-05), 2005.
7. T. Froehner, M. Nickles, G. Weiss. Towards Modeling the Social Layer of Emergent Knowledge Using Open Ontologies. In Proceedings of The ECAI-04 Workshop on Agent-Mediated Knowledge Management (AMKM-04), 2004.
8. J. Ullmann. Information Integration Using Logical Views. Proc. 6th Int'l Conference on Database Theory. Springer, 1997.
9. R. Giugno, Th. Lukasiewicz. P-SHOQ(d): A Probabilistic Extension of SHOQ(d) for Probabilistic Ontologies in the Semantic Web. In JELIA '02: Procs. of the European Conference on Logics in Artificial Intelligence. Springer, 2002.
10. F. Fischer, M. Nickles. Computational Opinions. Procs. of the 17th European Conference on Artificial Intelligence (ECAI'06), 2006. To appear.
11. B. Gaudou, A. Herzig, D. Longin, M. Nickles. A New Semantics for the FIPA Agent Communication Language based on Social Attitudes. Procs. of the 17th European Conference on Artificial Intelligence (ECAI'06), 2006. To appear.
12. R. Cobos. Mechanisms for the Crystallisation of Knowledge, a Proposal Using a Collaborative System. Ph.D. thesis. Universidad Autonoma de Madrid, 2003.
13. R. M. Cooke. Experts in Uncertainty: Opinion and Subjective Probability in Science. Oxford University Press, 1991.
14. I. Horrocks, P. F. Patel-Schneider. Reducing OWL entailment to Description Logic Satisfiability. Journal of Web Semantics, Vol. 1(4), 2004.
15. M. Bonifacio, P. Bouquet, R. Cuel. Knowledge Nodes: The Building Blocks of a Distributed Approach to Knowledge Management. Journal for Universal Computer Science, Vol. 8/6, 2002.
16. A. Farquhar, A. Dappert, R. Fikes, W. Pratt. Integrating Information Sources using Context Logic. In Procs. of the AAAI Spring Symposium on Information Gathering from Distributed Heterogeneous Environments, 1995.
17. J. Carroll, Ch. Bizer, P. Hayes, P. Stickler. Named Graphs, Provenance and Trust. In Procs. of the 14th International World Wide Web Conference, 2005.
18. A. Dragoni, P. Giorgini. Revisining Beliefs Received from Multiple Sources. In Frontiers in Belief Revision, H. Roth and M. Williams, Eds. Kluwer Academic Publisher, 431–444, 2001.