Article:

Karanasios, S, Allen, D, Norman, AWT et al. (3 more authors) (2013) Making Sense of Digital Traces: An Activity Theory Driven Ontological Approach. Journal of the American Society for Information Science and Technology, 64 (12). 2452 - 2467. ISSN 1532-2882

https://doi.org/10.1002/asi.22935

Published article:

Karanasios, S, Allen, D, Norman, AWT, Thakker, DA, Lau, L and Dimitrova, V (2013) *Making Sense of Digital Traces: An Activity Theory Driven Ontological Approach.* Journal of the American Society for Information Science and Technology, 64 (12). 2452 - 2467. ISSN 1532-2882

http://dx.doi.org/10.1002/asi.22935

Making Sense of Digital Traces: An Activity Theory Driven Ontological Approach

ABSTRACT

Social Web content such as blogs, videos and other user-generated content present a vast source of rich "digital-traces" of individuals' experiences. The use of digital traces to provide insight into human behavior remains underdeveloped. Recently, ontological approaches have been exploited for tagging and linking digital traces, with progress made in ontology models for well-defined domains. However, the process of conceptualization for ill-defined domains remains challenging; requiring interdisciplinary efforts to understand the main aspects and capture them in a computer processable form. The primary contribution of this paper is a theory driven approach to ontology development that supports semantic augmentation of digital traces. Specifically we argue that (1) activity theory can be used to develop more insightful conceptual models of ill-defined activities; which (2) can be used to inform the development of an ontology; and, (3) that this ontology can be used to guide the semantic augmentation of digital traces for making sense of phenomena. A case study on interpersonal communication is chosen to illustrate the applicability of the proposed multi-disciplinary approach. The benefits of the approach are illustrated through an example application, demonstrating how it may be used to assemble and make sense of digital traces.

1 Introduction

In recent years the social Web has emerged as one of the most interesting and challenging phenomena for researchers across several disciplines. The increasing numbers of journal special editions devoted to the social Web in both the social (Liang & Turban, 2011) and computer sciences (Boll, Jain, Luo, & Xu, 2011; Chen & Yang, 2011), multiple dedicated conference tracks and conferences, as well as the range of dedicated books (Page, 2012; Qualman, 2009) testify to the increasing importance of the topic.

Social media, or the social Web (Rheingold, 2000), is a group of internet-based applications that build on the ideological and technological foundations of the Web 2.0, and allow the creation and exchange of user-generated content (UGC) (Kaplan & Haenlein, 2010). Various published statistics and studies have emphasized the increase in use of the social Web by different demographic groups (Lenhart, Purcell, Smith, & Zickuhr, 2010). Mirroring the diverse range of users attracted to the social Web, there is also a breadth of content contributed and available (Agichtein, Castillo, Donato, Gionis, & Mishne, 2008). These "digital traces" manifest as videos and subsequent comments, blogs, stories and "Tweets", amongst other forms of UGC, as traces of social actions, interactions and transactions (Contractor, 2009), enabling people to document stories of their daily life and diverse experiences in unprecedented volume (Page, 2012). We see the potential of the social Web as a source of digital traces of real-life experiences, waiting to be analyzed and understood. However, while it has opened access to a wide range of primary data, the issues of organization and contextualization hinder the prospects of making sense of these data; making sense in terms of aggregating, organizing and linking digital traces from multiple sources to support browsing, search and understanding (in this way the term "making sense" is distinct from "sensemaking" as commonly used in information systems (Dervin, 1983) and management studies (Cornelissen, 2011)).

One area where the social Web provides a vast and largely untapped source of digital traces is in what we refer to as "ill-defined domains", such as interpersonal communication in corporate and multi-cultural settings. The terms "ill-defined", "unstructured" or "messy" domains refer to activities or problems with several workable solutions, depending on the context. Cognitively-complex processes which require human tacit knowledge are usually involved. Recently, decision science, artificial intelligence and software development (Lynch, Ashley, Aleven, & Pinkwart, 2006) have identified the challenges in dealing with ill-defined domains or problems. In particular, the need to identify and capture the knowledge required to operate in these domains.

To overcome the challenges to making sense of digital traces related to ill-defined activity, we argue that a combination of socio-theoretic techniques for developing insight and computer models for the power of automation is needed for the extraction and encoding of knowledge embedded in digital traces. Using an illustrative case of interpersonal communication we demonstrate a pioneering way by which activity theory can be used to develop conceptual models of ill-defined domains, which in turn are used to inform the development of an ontology. The ontology will guide the tagging of meaning in the collected content – a process called semantic augmentation. The annotated content can then be used for further reasoning.

The remainder of this paper is structured as follows. Section two reviews literature in the fields of information and computer sciences on the use of the social Web and making use of digital traces. It highlights the need for an integrated socio-technical approach to semantic augmentation of digital traces in ill-defined domains. Section three presents our activity theory driven approach for semantic augmentation. In Section four we describe our methodology to making sense of digital traces using an illustrative use case of an ill-defined activity. We focus on conceptualization of an activity, its conversion into ontology, subsequent validation and how we use it for semantic augmentation of digital traces. We also provide an example of how our approach may be used by presenting a platform that we have developed to facilitate browsing of digital traces. Section six discusses the implications of our work, identifies limitations and set priorities for future research. Section seven concludes the paper.

2 Literature Review

The purpose of this literature review is twofold. First, drawing upon both information science and computer science literature we examine the breadth of research on and methods of using and analysing digital traces, underscoring the rising interest in this field over the last decade. We also identify several limitations of the current research. Second, we argue that an alternate and new approach to making sense of digital traces, which draws upon computational methods in conjunction with theoretical frameworks which have been proven to help understand complex human behavior and activity is necessary to make sense of digital traces relevant to ill-defined domains.

2.1 Analysis and Use of Digital Traces in the Information and Computer Sciences

From the inception of the Web, the study of online content and information has been a topic of interdisciplinary endeavour (Turkle, 2005). Much of the early research focused on Web sites using "netnography" or "e-ethnography" (Hine, 2000), which focused on the analysis of Web site content, structure, links and changes over time (Schneider & Foot, 2005). Netnography studies also expanded beyond Web site analysis and explored discussions and interactions in the social Web, particularly in the fields of marketing and consumer behaviour, where access to large corpuses of data have offered in depth and difficult to access insights (Langer & Beckman, 2005; Nelson & Otnes, 2005). These approaches relied heavily on manual techniques for aggregating content and human interpretation for associating meaning to content.

Alongside the emergence of the social Web however there has been significant advancement in the tools available for synthesizing and analyzing digital data. The fields of infometrics and information retrieval in particular have drawn heavily on computational theories and tools in a nomothetic and positivistic attempt to assist humans in making sense of structured and unstructured data. The predominant approach used in these two traditions has been data mining to exploit large volumes of data in order to derive interesting patterns (Fayyad & Piatetsky-Shapiro, 1996). Similar approaches have also been extensively applied in Webometrics, a subfield of informetrics (Stock & Weber, 2006), which is concerned with the quantitative analysis of Web based information (such as hyperlinks). Recent work in these fields has been undertaken on blog entries (O'Leary, 2011), Twitter "tweets" (Thelwall, Buckley, & Paltoglou, 2011, 2012), Wiki data and most recently YouTube comments (Thelwall, Sud, & Vis, 2012). Despite these growing efforts, work to apply informetric analysis of digital traces and the methods for organizing and providing access to UGC data are still relatively new (Efron, 2011). Contextual information, on which the interpretations are based, plays a crucial role for the analysis of experience captured in these digital traces.

By far, sentiment analysis and prediction remain the largest area of application in the use of digital traces and have been for some time (Nasukawa & Yi, 2003). These approaches have been used to predict real-world outcomes, such as sales (Asur & Huberman, 2010), mood and market performance (Bollen, Mao, & Zeng, 2010) based on Twitter "chatter". Two areas of concern have, however, been identified. Thelwall (2006) has argued that the highly positivist, quantitative frameworks used by many studies are problematic and that there is a need for complementary qualitative methods in order to fully understand the data (Thelwall, 2006; Thelwall & Vaughan, 2005). The degree to which on-line behavior is related to offline phenomena and characteristics has also been questioned (Gayo-Avello, 2012; Park & Thelwall, 2008).

Emergent research on the use of semantics to make sense of digital traces has linked social tags to a controlled vocabulary (Yi, 2010) and uses tag analysis to uncover semantic relations (Yoon, 2012). Research within this area has also focused on taking this knowledge into systems design through, for example, the integration of "folksonomy" tags (collaborative methods of creating and managing tags to annotate and categorize social Web content) into a taxonomy to develop systems within a corporate environment (Tsui, Wang, Cheung, & Lau, 2010). Indeed, the semantic Web offers new approaches for making sense of digital traces utilizing semantics of data and domain to derive interesting patterns and connections. It has been used in a range of ways from profiling users on the social Web for adaptation (Abel, Gao, Houben, & Tao, 2011), profiling online communities (Rowe, Alani, Angeletou, & Fernandez, 2012), archiving social Web content for future use (Edelstein et al., 2011) and mining the digital traces to represent "collective intelligence" (Diplaris et al., 2011; Noh, Park, Park, & Lee, 2010). In the case of collective intelligence, Diplaris et al., (2011) and Noh et al.,

(2010) propose collective knowledge systems, which combine social Web content and semantic Web technology to process the emergent knowledge from the social Web. Collective knowledge systems enable computation and inference of the collected information, leading to answers and discoveries that are not found in the human contributions (Noh et al., 2010). Such new discoveries and answers are referred to by the authors as “emergent knowledge” (Gruber, 2007) and can be utilized to design powerful applications, such as for disaster relief (Gao, Barbier, & Goolsby., 2011). Noh et al., (2010) focused on finding and accessing emergent knowledge from real-world blog postings. They considered blog postings as experiences of people on particular topics. They consider the user contributions in social Web (in their case blogs) as expressions of various experiences of the users (i.e. digital traces of real-world experience). The emphasis on generating emergent knowledge from the digital traces in their work is in line with our premise on using semantic technologies for making sense of digital traces.

While it was not the purpose of this literature review to provide an exhaustive discussion of the current and ever expanding use of digital traces, it is unequivocal that empirical investigation concerning how to make use of digital traces effectively and efficiently remains a relatively new and broad area of scholarship. Effectively, in terms of identifying the relevant concepts and relationships to be exploited in the digital traces; efficiently in terms of the algorithmic approach to process large volumes of content; and broad in terms of the scope of application and methods to be applied. We also note two other limitations of the current approaches, which present opportunities and challenges for research. The first is that the use of underlying socio-theoretical frameworks, such as those employed in social science research to examine and understand real-world settings that can help to understand complex human behavior and activity is lacking from the study of digital traces. The second is that the use and application of digital traces in the area of ill-defined activities and messy domains remains under studied.

2.2 An Integrated Approach to Making Sense Ill-Defined Domains

Building on the previous section, here we argue that semantic augmentation of digital traces is one way of approaching the problem of making sense of digital traces of an ill-defined activity. We also argue that to inform semantic augmentation a relevant ontology is required. However current ontologies are mostly developed for well-defined domains. It remains challenging to develop conceptual models of ill-defined domains for the knowledge engineering process. We propose that activity theory can be used to inform the development of an ontology for ill-defined activities and discuss its use for both studying messy domains and informing systems design and ontology development.

2.3 Semantic Augmentation

Semantic augmentation is a process of attaching semantics (in the form of "concepts") to a selected part of a text to assist automatic interpretation of the meaning conveyed by the text. It is being increasingly adopted as the main mechanism to aggregate and organize social Web content by linking it with ontology concepts. This has primarily taken place in well-defined domains, where researchers have moved beyond traditional document-level meta-data approaches and started to investigate factual content in text to support more intelligent and semantic-based access (Cui, 2012). This approach relies on tagging; which assigns a set of concepts to resources as a way of providing meaning. Semantic units are extracted mainly by identifying named entities (e.g. people, organizations, locations). This has become a powerful way of organizing, browsing, and publicly sharing personal collections of resources on the Web (Berners-Lee, Hendler, & Lassila, 2001; Kim, Breslin, Kim, & Choi, 2010; Noh, et al., 2010).

In order to semantically augment digital traces it is necessary to have a semantic model in the form of an ontology (Bontcheva & Cunningham, 2011). Ontologies are an explicit specification of a conceptualization (Gruber, 1993) and model a mini-world of the task domain in a computerunderstandable way and are, therefore, useful for improving discovery and access to digital content (Greenberg & Méndez, 2009; Kim, 2005).

2.4 Ontology Development

Computer scientists and knowledge engineers have made progress in creating ontologies for well-defined domains, which are defined as domains where there exists a systematic way to determine when a proposed solution is acceptable (Minsky, 1995); with many successful implementations of large-scale ontologies in well-structured subject areas e.g. in the biomedical/biological (Cui, 2010; Harris & Clark JI, 2006), navigation (Kuhn, 2001) and medical (Rector et al., 1998) domains. With the recent proven success of the semantic Web and ontologies (Bizer, Heath, & Berners-Lee, 2009), the field is ready to take on the challenges offered by ill-defined domains, and better understand the key issues in the benefits of semantics for applications in these domains (e.g. for reasoning, aggregation, automation). Along these lines, Kim noted (2005 p.1175) that “the domains which contain the most valuable knowledge, such as the medical and the business sectors, are prime areas for future ontology applications”. One of the major challenges remains developing conceptual models of ill-defined domains in order to inform the knowledge engineering process. A reason for this is that the conceptual representation of such domains involves domain experts, study subjects and understanding of complex and fluid human activities, as well as theoretical frameworks to guide conceptualization. Activity theory is one theoretical framework that has been used to inform ontology development (Kuhn, 2001; O'Leary, 2010) and to understand and represent ill-defined activities (Authors, 2013; Wells, 2002).

2.5 Informing Ontology Development Using Activity Theory

Here we focus on the two major theoretical contributions that activity theory offers to modelling ill-defined activity and ontology development. Detailed description of activity theory, comparison to other contemporary social theories and theoretical arguments surrounding the concepts can be viewed elsewhere (cf. Foot, 2002; Kaptelinin, 2005). The first contribution centres on the conceptualization offered by the activity system structure for understanding and framing complex activity. This structure provides a "root model" of human activity (Engeström, 1987) and a conceptual framework for inquiry. In addition, there is similarity between the activity system structure and the ontological structure (i.e. "concepts" and "relationships between concepts") which lends itself to bridging the transition between the human based and the machine based models. The second contribution is the variable level of granularity for analyzing activity (in terms of "actions" and "operations"). This provides a flexible theoretical lens for zooming in and out of an observed sequence of interactions to identifying the embedded knowledge. These aspects of activity theory are elaborated in the following paragraphs. Therefore in this work we engage the activity system concepts, rather than broader activity theory as a theory based framework.

An activity system perspective posits that the essential site for analyzing interaction between actors and collective structures is the "activity". An activity is undertaken by a "subject", which can be either an individual or collective (organizations, a team etc.), with an underlying "motivation(s)" to act upon an "object" (typically a person, collective or thing) in order to achieve an "outcome". Objects have a life of their own, which is emphasized when objects resist the attempts of the subject to control them. This is particularly true when it is other people that are the object of activity (Engeström, 2000), such as in the case of dyadic communication (Wells, 2002).

The achievement of the object arises through the use of physical (such as IT, hammer etc.) and mental "tools" (such as signs, language, experience), which follows the principle that humans on most occasions interpose an artefact between themselves and the object of interest, thereby enabling them to act more effectively (Wells, 2002). In addition, an activity includes a broader "community", who are stakeholders and collaborators in it. An activity is underpinned by cultural-historical rules and norms which are associated with values and social judgements which influence the activity and subject to a “division of labor” which concerns the accomplishment of the task that contribute to the attainment of the object (these are shown in Figure 1).

(Insert figure 1 here)

An important distinction in activity theory is between “activity”, “action” and “operation”. This posits that humans engage in goal-orientated actions which individually do not necessarily directly contribute to the attainment of the object of an activity (Leont'ev, 1978). That is, actions are only accomplished because they realise a specific activity (Roth, 2007). However, collectively the actions (undertaken by individuals or groups) lead to the satisfaction of the motive through realization of the object and typically only make sense in the context of a shared work activity. Beneath the level of actions, are operations which may be conscious when initially learned but may become subconscious or routinized (cf. Authors 2011 for detailed example). Therefore, an activity includes a number of actions, and each action may constitute many operations. A widely used example is the building of a house (the activity), which involves building the foundations, fixing the roof, etc (the actions), and using a hammer, laying the bricks, etc (the operations).

The three levels of activity are not fixed. Activities that exist at one time may become actions at a later time, likewise an action can become an operation (Bødker, 1991; Leont'ev, 1978). The fluidity between activity-actions-operations allows researchers to understand an activity, while recognising that an activity may become an action and an action an operation, and vice-versa (Authors 2011). A benefit of this distinction and fluidity between activity-action-operation is that it can also be observed in the ontological structure.

Activity theory has been employed in several fields of scholarship for understanding complex human activity (Engeström, 2000; Hassan Ibrahim & Allen, 2012; Wells, 2002) and informing system design (Chen, Sharman, Rao, & Upadhyaya, 2013; Korpela, Mursu, & Soriyan, 2002). Demonstrating its usefulness in ontology development Kuhn (2001) used some of the hierarchical concepts available from activity theory, namely activity-actions-operations, to structure an ontology. O'Leary (2010) used an activity theory driven approach as a template to capture the context of individual activity in an organization (an "enterprise ontology").

Despite the use of activity theory for studying both complex and ill-defined activity, and ontology development, its use for ontology development of ill-defined activity remains relatively unexplored, rudimentary and lacking actual application. We argue that the insights offered by activity theory through the articulated contributions can better support the development of an ontology related to ill-defined activity.

3 Research Question

The foregoing discussion highlighted three key arguments, drawn from the literature. First that semantic augmentation of digital traces can be used to make sense of digital traces. Second, that to inform semantic augmentation of digital traces a relevant semantic model in the form of an ontology is required. Third, that activity theory can be used to inform the development of an ontology for an ill-defined activity. Based on these interrelated arguments the research question that forms the basis of this paper is:

How does an activity theory driven approach to ontology development and semantic augmentation support making sense of digital traces in ill-defined domains?

To answer this research question in the next sections we outline our methodology for modelling ill-defined activity, ontology development and semantic augmentation of digital traces, and demonstrate the utility of this approach through an illustrative case. Alongside this research question, we also explore the related issue of the potential uses of digital traces for theory building, expanding knowledge and influencing the design of intelligent systems, amongst other possibilities which require good understanding of human behavior; by doing so, we extend the research on the use of the social Web.

4 Activity Theory Driven Approach for Semantic Augmentation

To address the research question, we separate our methodology into three distinct phases, as illustrated in Figure 2: (1) using activity theory to develop a conceptual model of ill-defined domains; (2) converting the conceptual model into an ontology; and, (3) applying the ontology and semantics to augment digital traces for understanding and reuse.

(Insert figure 2 here)

The first phase consists of using the concepts available through activity theory to capture and conceptualize an ill-defined activity. To do this social science data collection and analysis methods are applied in order to develop a rich conceptualized model of the activity. The second phase involves converting the conceptual model into logical form as an ontology. In the third phase, a semantic augmentation service is used to automatically tag the digital traces, using the ontology. The following section introduces an illustrative case study followed by details on the application of our methodology in the case study.

5 An Illustrative Case Study

The illustrative case we have selected is that of interpersonal communication, defined as communication involving message transaction (sending and receiving messages) between two or more dependent participants (West & Turner, 2011). Interpersonal communication is an area where there is a large volume of digital traces (e.g. individuals blogging on their experience in a job interview, Tweeting about their encounters in a foreign country etc.) and continued interest amongst researchers, business and individuals in developing better understanding and skills. Within this illustrative case we focus on the specific setting of the "job interview". As there are a number of formats for a job interview (for instance, group interview, panel interview, telephone interview, one-on-one interview etc), we narrow our activity to the dyadic interview, conducted in a face-to-face setting. The job interview falls under the umbrella of ill-defined as there are several complex personal, organizational, non-verbal and cultural factors that influence its conduct (Goldberg, 2003; Millar & Tracy, 1997; Raffler-Engel, 1983). We envisage that the semantic augmentation of digital traces related to a job interview case could be of use to learners (job interviewers or job applicants), trainers (human resource departments) and also social scientists for making sense of large corpuses of digital traces for analysis, learning and theory building, amongst other potential uses that would benefit from the synthesis of multiple sources and viewpoints.

5.1 Activity Analysis and Conceptualization

In order to model the job interview, data were collected from a number of main data sources using qualitative research techniques. This included interviews with key informants and interviewees, interviewers, trainers, a review of organizational documents and attendance at job interview workshops in order to investigate the key concepts of the activity. Focusing on several sources provided triangulation and a more holistic perspective whilst limiting the potential bias from relying on a single data source (Eisenhardt, 1989). Full details of the data collection and the methods applied are explained elsewhere (cf. Authors 2011). Figure 3 provides a graphical representation of the resulting activity model for the job interview activity. Here we only illustrate the upper-most level of abstraction. The associated actions and operations, which are intricate in detail, can be viewed here[i].

(Insert figure 3 here)

The conceptualization of the activity in Figure 3 illustrates that an interviewer (subject) undertakes a job interview (the activity), motivated by the desire to assess the applicant (the motivation). The interviewer interacts with the applicant (the object), using physical and mental tools. While the physical tools in this case (i.e. pen, paper, furniture, clothing and so on) are somewhat obvious, the use of the mental tools including language and interpersonal skills, are less well-defined in terms of when and how they may be employed to interact with the object. The job interview activity is also bounded by several interrelated rules and norms which are influenced by the organization, the cultural dimensions (e.g. national/cultural/social differences between individuals), the contextual setting (e.g. corporate, educational environment), social values (i.e. display rules and etiquette), legal requirements (e.g. pertaining to questioning) and activity norms (i.e. the "typical" process of performing a job interview). The activity is influenced by a relatively narrow community and follows a structured and commonly understood division of labor, which also increases the pressure and anxiety within the activity. Such a model provides a base conceptual representation of the activity; capturing its complexity in an abstract form.

5.2 Logical Encoding: Transferring Conceptual Model into Ontology

The process of transformation from conceptual form to logical form involved a combination of knowledge engineers and social scientists developing a knowledge glossary using the concepts and relationships identified in Figure 3 and a Controlled Natural Language (CNL) based authoring tool (cf. Bao, Smart, Braines, & Shadbolt, 2009), ROO[ii] (cf. Authors, 2011) to transpose the conceptual model into an "activity model ontology" (hereafter, referred to as AMOn). ROO was then used to translate the CNL into the corresponding Description Logic (DL) (cf. Baader & Nutt, 2003) needed for machine reasoning.

A technical challenge faced in the knowledge engineering process was the scalability of the ontology. To enable the evolution of AMOn it was developed as a group of smaller but interlinked ontologies, following the principle of modularization (Parent & Spaccapietra, 2009). This provided the advantage of re-using parts of the ontology, hence allowing more effective management of complexity and promoting understanding of the ontologies across different contexts, facilitating expansion and scalability. Similar to Almeida and Barbosa (2009), we follow a 3-layered model which consists of an upper abstract layer, middle re-usable layer and lower case-specific layer. Each layer can be further developed as one or more ontology modules. Figure 4 outlines the three main layers in AMOn.

(Insert figure 4 here)

The upper abstract layer covers the base concepts provided by activity theory such as: Activity, Tools, Action, Operation, Motivation, Outcome, Subject and Rules and Norms (see Figure 5). Relationships amongst these concepts are also defined. For example, the concept of Tool is defined as "*something that is used by Subject in an Activity to achieve an Object*". This layer is the root model of human activity and the foundation of our understanding of the job interview activity. Table 1 outlines some of these concepts from the upper abstract layer with CNL and the corresponding DL representation.

(Insert figure 5 here)

(Insert table 1 here)

In the middle layer, we define concepts that are relevant to a specific domain, in our case, interpersonal communication. Although the concepts in this layer were initially derived for the job interview activity, they are generic enough for reuse in other activities related to interpersonal communication in work settings. Figure 6 presents several of the concepts which are derived from the conceptual activity model. Table 2 outlines some of these concepts from the middle layer with CNL and its corresponding DL representation.

(Insert figure 6 here)

(Insert table 2 here)

Our modelling approach allows us to use relevant external ontologies to enhance the coverage of the concepts in the AMOn modules. This is particularly useful to improve the coverage of expansive concepts, such as in the case of concepts related to Body Language, where we drew on an external ontology[iii] (cf. Authors, 2011) (see Figure 7).

(Insert figure 7 here)

We also reused SentiWordNet[iv] and WN-Affect[v] ontologies to improve the coverage of Emotion. A link is also made between these two sets of concepts ("Body Language" and "Emotion") as emotions are usually expressed through body language. We also link to the FOAF (Friend of a Friend)[vi] ontology for its description of the concept of the "Person" to further specify the concept of the Subject in an activity.

The third, case-specific, layer encapsulates the concepts that emerged from analysis of job interviews. These concepts expand from the interpersonal communication concepts in the middle layer. For example, the specialization of the concept "Mental Tool" in the middle layer is expanded to encapsulate Mental Tools explicitly relevant to the job interview such as interviewing skill, experience and so on.

Table 3 presents a summary of the ontological features of AMOn in terms of size (number of classes, properties and instances/objects), expressivity (conforming to SHOIQ in DL (Horrocks & Sattler, 2005)) and complexity of the core knowledge (captured by axioms)[vii].

(Insert table 3 here)

5.3 Validation of AMOn

Before proceeding to using the ontology for semantic augmentation we performed an evaluation to answers two key questions: (1) is the breadth of concepts captured by AMOn representative of the job interview case? and, (2) which theoretical concepts suffered in the transposition from conceptual model to AMOn? These questions served the purpose of validating the ontology before it was used to semantically augment digital traces and illuminate the transition of activity theory concepts from conceptual to computerized form.

Whilst various approaches to evaluating an ontology exist, "competency questions" remains the most common approach (Almeida & Barbosa, 2009; Uschold & Grüninger, 1996). This approach stipulates that an ontology must be able to represent the competency questions using its terminology and answer these questions using the axioms (Gruninger & Fox, 1994). Our evaluation started with eliciting written competency questions from five human resource managers and trainers with experience in undertaking job interviews. In total 78 competency questions were received. The second step was to extract keywords from the competency questions using a process of thematic analysis and mapping keywords to the conceptual model and AMOn respectively. A total of 143 keywords were extracted. During mapping, we analyzed whether a match was "exact", "synonymous" (as derived from WordNet or Oxford Dictionary) or at an "abstract level". The analysis was undertaken jointly by a social scientist and knowledge engineer. Table 4 provides an overview of the analysis.

(Insert figure 4 here)

As shown in Table 4, out of 143 keywords extracted from the competency questions, an exact match was found with 26 concepts from the activity model (hence also in AMOn), and a further 4 concepts drawn from external ontologies. Therefore, a total of 30 keywords found an exact match in AMOn. Similarly, 34 synonyms of the keywords were covered by the activity model with a further six brought in by the external ontologies. Hence, 70 keywords were covered by concepts in AMOn, exactly or synonymously. Mapping at the abstract level, 96 keywords found an exact match in AMOn with concepts drawn from the activity model; with a further 7 concepts added by bringing in external ontologies. Overall, 103 out of the 143 concepts were covered by AMOn at the abstract level, which offered a 72 percent coverage.

On closer examination of the unmatched keywords, some were related to physical artefacts expressed in the competency questions (such as “refreshments”). The analysis of the breadth of concepts was also used to illuminate the transposition from activity theory concepts to the ontology. The analysis revealed some weakness in accounting for concepts related to the individual characteristics, such as religion, gender, aggressive, shy and so on, of the “subject” (the interviewer) and “object” (the interviewee) in the conceptual activity model. This limitation subsequently propagated into AMOn. Another important finding was that the fluidity afforded by the conceptual model is not entirely supported in AMOn. For instance, the terms “rules and norms” and “division of labor” are difficult to exactly articulate as required in an ontology. At present, such concepts are described as atomic concepts but do not include deeper semantics. For example, the concept “division of labor” is not informative, however, adding axioms that specify it further makes it more useful for reasoning processes; for example, “division between whom”, “what labor means”, “what are the types of labor”. More broadly, some of the concepts within activity theory such as “object” remain disputed (e.g. as capturing the "true" motive (Kaptelinin, 2005)). In order to allow for precise articulation of the concept of object, it was referred to in AMOn as "a group, individual or physical thing", which does not maintain the full ontological commitments to activity theory, but is a balance that is necessary in order to follow the guidelines of ontology development. In summary, the validation exercise identified some examples in which activity theory concepts suffered in transition to a logical form, resulting in some of the more nebulous activity theory concepts being constrained to a more atomic level.

To conclude, the evaluation exercise revealed that AMOn was fit for purpose with respect to the representation of interpersonal communication for the job interview case.

5.4 Semantic Augmentation

In order to aggregate digital traces, semantic augmentation was used as a means to link digital traces to the concepts in AMOn. This process indirectly incorporates the contextual knowledge about the job interview, as a result of the application of activity theory. Figure 8 illustrates the architecture of the semantic augmentation service which was developed to automate the process.

(Insert figure 8 here)

The service takes textual content as input and produces semantically augmented content. It has three main components: Information Extraction, Semantic Indexer and Semantic Repository. The Information Extraction component takes in the textual content and produces annotated sets of extracted entities and other metadata (e.g. ontology Universal Resource Identifier [URI] and type information). The Semantic Indexer component draws on the Semantic Repository and converts the annotated sets into Resource Description Framework (RDF) triples (subject-predicate-object) i.e. the semantically augmented content. The Semantic Repository combines the functionality of an RDF-based Database Management System and an inference engine. It can store data and evaluate queries, guided by the ontologies and metadata schemata.

Regarding the implementation, OWLIM[viii] was used for the Semantic Repository. The Information Extraction component used General Architecture for Text Engineering (GATE)[ix] and the Semantic Indexer component was built using the Sesame SPARQL API[x] and extensively utilized SPARQL queries.

The augmentation process consists of two stages: (1) set up (performed offline manually); and, (2) processing (performed on-the-fly automatically).

(1) Set up: This includes selection of ontologies for the domain (in this case AMOn). These ontologies are used to build a list of known entities (called Gazetteers - a format used by the Information Extraction engine). The mapping from ontologies to gazetteers exploits the label properties of ontology standards such as RDF Schema (RDFS) and Simple Knowledge Organization System (SKOS) e.g. rdfs:label, skos:prefLabel, skos:altLabel. A list of gazetteers are prepared for each class (e.g. rdfs:class and owl:class) from the ontologies. Rules are defined for pre-processing of the content and for post-processing of the annotation sets from Information Extraction. JAPE grammar is used for the definitions.

(2) Processing: The input to the processing algorithm is textual content. The pre-processing JAPE grammar is applied on the textual content to identify and extract the part of text for augmentation. The linguistic processing techniques (such as sentence detection and splitting, tokenisation, part-of-speech tagging, sentence chunking and word stemming) are then applied on the textual content. This process produces a surface form (i.e. linguistically annotated text with nouns, verbs, adverbs and adjectives). The gazetteer component is applied on these surface forms which matches the ontological labels to the surface forms and attaches an ontology concept URI to the surface forms. Then in the next step, the Semantic Indexer converts the annotation sets in semantic format (i.e. set of triples), checks the existing index for the content and stores new or updated indexes into the Semantic Repository[xi].

5.5 An Application to Make Sense of Digital Traces on Interpersonal Communication

The output of our integrated approach described in this paper could be used for several applications that involve making sense of ill-defined activities. An Intelligent Content Assembly Workbench (I-CAW)[xii] platform was developed to demonstrate the value added by our approach – using the job interview setting as an example domain.

To illustrate how I-CAW could be used to make sense of real-life experiences from the digital traces, let us consider the following learning scenario.

Jane is new to an organization and is commencing her first job in a human resources department. Her new role will involve interviewing applicants for a vacant position. After inquiring she was advised by some colleagues that interpersonal communication aspects such as non-verbal behavior can play a crucial role in the conduct of an interview and it is important that she is aware of them to help her successfully perform the interview. She wants to learn more about this. She knows that there is a large volume of digital traces (videos and comments, blogs, stories etc) on the Web but it is time consuming to search this content and create meaningful associations between the content and draw inferences. Using I-CAW she searches the term "body language" and is offered several digital traces in the form of videos on YouTube (and comments) on job interviews, people's real experiences represented in blogs, amongst other forms. This hybridization of heterogeneous content is made possible with the semantic augmentation service, which extracts key terms from these different textual contents and links them to ontological concepts, stores them in the Semantic Repository and makes them available for querying.

In I-CAW Jane is able to browse through the recommended videos, comments and other digital traces. While reading about a particular blog, she learns that eye-contact can be important and that there may be several possible interpretations. She clicks on eye-contact and is particularly interested in the link pointing to "Body Language Meanings" (derived from the ontology by I-CAW), which leads to a list of manifestations of eye-contact (gaze, stare, and its possible interpretations as meaning anxiousness or how it can mediate a conversation etc). From there, she clicks on “Nervousness” and arrives at another page with a collection of related YouTube videos, personal stories and comments. From the information she discovers other types of body language signals (other than eye-contact) that may indicate nervousness (derived from the relationships defined in AMOn). During this process of inquiry, Jane learns from the diverse comments about the different interpretations of people from different cultures and the different perspectives (e.g. interviewer vs. job applicant, novice vs. experienced individuals).

In this scenario I-CAW has intelligently coalesced and made sense of different digital traces by exploiting the ontology and semantic augmentation. Users can develop insights into the phenomenon based on previous experience and insights of others as captured by the digital traces.

Figure 9 illustrates how I-CAW makes sense of digital traces according to the user’s interest. In this Figure a user browsed interpersonal communication aspects related to the job interview and selected a particular comment of interest. From the content of the comment, I-CAW pulled together a collection of other relevant content, including a video from YouTube with a comment which suggested certain indicators of nervousness observed in the video. From here the user may click on nervousness or another keyword for other content.

(Insert figure 9 here)

6 Discussion

The extant research on the use and application of digital traces highlights various streams and approaches. We argued that one of the shortcomings of the current approaches was that theoretical frameworks which have aided in the understanding of ill-defined domains in the social sciences have been neglected in making sense of digital traces. We proposed our approach, based on activity theory and drawing on computational methods of ontology and semantic augmentation. In order to demonstrate the value of this approach this paper sought to answer the research question how does an activity theory driven approach to ontology development and semantic augmentation support making sense of digital traces in ill-defined domains?

The foregoing sections outlined our steps in developing and converting an activity theory founded conceptual model into an ontology for the purpose of guiding the semantic augmentation of digital traces. The evaluation of the activity model and AMOn showed that activity theory provides the strong conceptual tools to represent an activity and that AMOn builds on the resulting conceptual model, providing a logical model of activity, which also drew on external ontologies. The work therefore, presents a step-change, one that builds on netnographic approaches, but is distinct from automated approaches of synthesizing and analyzing digital traces. We have argued that this is necessary when dealing with the domain on which this research centres, which is messy and requires the support of social theory to understand. However, while activity theory is useful for understanding complex human activity, and provides a structure and concepts which are commensurable with ontology development (in particular the structure) it must be emphasized that, as expected, several theoretical concepts suffered in the transition to logical form. For instance, some of the fluidity in concepts such as “division of labor” was found to be overly dynamic for succinct transposition into logical form. Likewise, while the notion of the “rules and norms” of an activity can be observed and captured by social-scientists they are not amenable to the less flexible conditions of ontology development. Such limitations meant that some concepts were described as atomic concepts and did not include deeper semantics. This is a significant finding, lesson and challenge; one that we emphasize, rather than presenting a straightforward and uncomplicated transition from contemporary social theoretical frameworks to ontology. Others have noted that some activity theoretical concepts would suffer in their transposition into logical form and although “*such creative transformations are legitimate, possible, desirable and probable*” (Baker, 2000 p. 129).

While the application presented remains suggestive it demonstrates the power and novelty of our work. Our work brings together recent developments in the fields of information science and computer science that have separately examined ways to aggregate, make sense of, and use, the social Web content. In particular it sets itself apart from approaches that have drawn upon data mining by utilizing semantics and activity theory as a foundation for understanding ill-defined activity. The development of I-CAW to visualize our semantic augmentation adds to increasing interest in methods of visualizing semantic data (Bizer, et al., 2009; Dadzie & Rowe, 2011).

6.1 Limitations and future research

As with any new research our work contains several limitations, which also present interesting directions for future research. One particular area that requires further investigation concerns the use of digital traces and their validity as “real” representations that can be used to develop understanding of human behavior. Whilst there has been suggestion that the use of social Web can be used for prediction, sentiment analysis and understanding of social-phenomenon (Asur & Huberman, 2010; Thelwall, et al., 2011), there has been counter arguments as to its validity for these purposes (Gayo-Avello, 2012; Park & Thelwall, 2008). This is particularly true for our approach, where we have considered digital traces as valid representations of the real-world. This is an interesting and much-needed avenue for inquiry as researchers and citizens place more emphasis on social Web data for understanding phenomena.

We are also alert to the dependence of our approach on a large corpus of digital traces in order to be useful. In the case of the job interview, the volume of digital traces was extensive; however, other specific scenarios may not be heavily represented in digital traces. While we have validated the ontology, an area for inquiry is the extent to which the activity model, and hence AMOn, is generalizable to other representations of job interview. For instance, our conceptualization of the job interview was derived from the UK work setting, it would be interesting to understand how generalizable this is and how adaptable AMOn is to other job interview representations that are captured in digital traces.

The case presented within this study focuses on a small-scale case study of the job interview. It was not the intention to provide significant insights into the job interview; rather it acted as our experimental setting. As such some activity theory concepts have not been explored. For instance, the notion of cultural-historical activity and contradictions, which have been the emphasis of our previous works (Authors, *In press*; Authors, 2013) was not considered because of the difficulty in capturing these in logical form and because it was beyond the exposition of the ontology application. However, within the boundaries of the work we have maintained ontological and epistemological commitments to activity theory.

In our next phase of research we will apply our approach to a further two ill-defined interpersonal communication scenarios. These include a foreign visitor mentoring programme and a doctor-patient interaction. The purpose will be to apply our methodology to another setting to determine the generalizability of our approach, expand the ontology and determine which levels of modularization are re-usable. A further objective will be to learn from the different sources of digital traces available and evaluate I-CAW as a learning tool, something that we have only briefly explored in this paper. This will provide evidence concerning the utility of digital traces for learning. Another direction for future research is developing appropriate techniques for capturing, aggregating and comparing viewpoints of digital traces; some of our ongoing has begun to address these issues (Authors 2012; Authors, 2012).

7 Conclusion

The work described in this paper presented an integrated approach, drawing upon socio-theoretical foundations, to deliver significant synergies in making sense of digital traces. As the use of the social Web increases this is likely to be an increasingly rich area for development, both in the sense that its richness and value as a resource will increase rapidly and also in the sense that we will have a greater expectation of being able to make use of such rich corpuses of data. In this paper, we argued, using an illustrative case on interpersonal communication, that activity theory can be used to develop more insightful conceptual models of ill-defined activities, which can be used to inform the development of an ontology. The resulting ontology was then used to guide the semantic augmentation of digital traces for making sense of phenomena. We presented I-CAW and a suggestive example of how it may be used to help an individual make sense of interpersonal communication in the job interview setting. The case we have demonstrated here bridges the structured tools of ontology development and the unstructured worlds of ill-defined domains making use of activity theory to provide a means of ontology development and semantic augmentation, which, in turn, render the hitherto hard-to-capture-and-aggregate digital traces both relevant and accessible for several suggestive purposes. We do not claim that this is a finished piece of work, rather, that it opens up possibilities on which we and others can build.

Acknowledgement

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no ICT 257831 (ImREAL project).

References

Authors, (In press) in *MIS Quarterly*

Authors, (2013) in *Information Systems Journal*

Authors, (2011) in *Journal of the American Society for Information Science and Technology*

Authors, (2011) in *19th User Modeling, Adaptation and Personalization (UMAP)*

Authors, (2011) in *Journal of Web Semantics*

Authors, (2011) in *E-challenges*

Abel, F., Gao, Q., Houben, G.-J., & Tao, K. (2011, May 29-June 2). *Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web.* Proceedings of the The Semantic Web:

Research and Applications, 8th Extended Semantic Web Conference, (pp. 375-389) Heraklion.

Agichtein, E., Castillo, C., Donato, D., Gionis, A., & Mishne, G. (2008). *Finding high-quality content in social media.* Proceedings of the Proceedings of the international conference on Web search and web data mining, (pp. 183-194) Palo Alto: ACM.

Almeida, M. B., & Barbosa, R. R. (2009). Ontologies in knowledge management support: A case study. *Journal of the American Society for Information Science and Technology, 60*(10), 2032-2047. doi: 10.1002/asi.21120

Asur, S., & Huberman, B. A. (2010, Aug 31- Sept 3). *Predicting the Future with Social Media.* Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), (pp. 492-499) Toronto: IEEE Computer Society.

Baader, F., & Nutt, W. (2003). Basic description logics. In B. Franz, C. Diego, L. M. Deborah, N. Daniele & F. P.-S. Peter (Eds.), *The description logic handbook* (pp. 43-95): Cambridge University Press.

Baker, M. (2000). International Journal of Artificial Intelligence in Education. *The roles of models in Artificial Intelligence and Education research: a prospective view, 11*(2), 122-143.

Bao, J., Smart, P., Braines, D., & Shadbolt, N. (2009). *A Controlled Natural Language Interface for Semantic Media Wiki Using the Rabbit Language*. Paper presented at the Workshop on Controlled Natural Language (CNL'09), Marettimo Island.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The Semantic Web. *Scientific American, 284*(5), 34-43.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data - the story so far. *International Journal on Semantic Web and Information Systems, 5*(3), 1-22.

Bødker, S. (1991). *Through the Interface. A Human Activity Approach to User Interface Design*. Mahwah: Lawrence Erlbaum Associates Inc.

Boll, S., Jain, R., Luo, J., & Xu, D. (2011). Introduction to special issue on social media. *ACM Trans. Multimedia Comput. Commun. Appl., 7S*(1), 1-2. doi: 10.1145/2037676.2037682

Bollen, J., Mao, H., & Zeng, X.-J. (2010). Twitter mood predicts the stock market. *Journal of Computational Science, 2*(1), 1-8.

Bontcheva, K., & Cunningham, H. (2011). Semantic annotation and retrieval: Manual, semi-automatic and automatic generation. In D. F. J. Domingue, and J. A. Hendler, (Ed.), *Semantics of Social Media Streams book of Semantic Web Technologies*: Springer.

Chen, H., & Yang, C. C. (2011). Special Issue on Social Media Analytics: Understanding the Pulse of the Society. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 41*(5), 826-827.

Chen, R., Sharman, R., Rao, H. R., & Upadhyaya, S. J. (2013). Data Model Development for Fire Related Extreme Events: An Activity Theory Approach. *MIS Quarterly, in press*.

Contractor, N. (2009). *Digital Traces: An Exploratorium for Understanding and Enabling Social Networks*. Paper presented at the Symposium: Interdisciplinary Approaches to the Study of Large-Scale Human Networks. Annual meeting of the America Association for the Advancement of Scienec (AAAS), Columbus.

Cornelissen, J. P. (2011). Sensemaking Under Pressure: The Influence of Professional Roles and Social Accountability on the Creation of Sense. *Organization Science 23*(1), 118-137.

Cui, H. (2010). Competency evaluation of plant character ontologies against domain literature. *Journal of the American Society for Information Science and Technology, 61*(6), 1144-1165. doi: 10.1002/asi.21325

Cui, H. (2012). CharaParser for fine-grained semantic annotation of organism morphological descriptions. *Journal of the American Society for Information Science and Technology, 63*(4), 738–754. doi: 10.1002/asi.22618

Dadzie, A.-S., & Rowe, M. (2011). Approaches to Visualising Linked Data: A Survey. *Semantic Web, 2*(2), 89-124.

Dervin, B. (1983, May). *An overview of sense-making research: concepts, methods, and results to date.* Proceedings of the Paper presented at annual meeting of the International Communication Association, Dallas, Texas.

Diplaris, S., Sonnenbichler, A., Kaczanowski, T., Mylonas, P., Scherp, A., Janik, M., & Papadopoulos, S. (2011). Emerging, Collective Intelligence for Personal, Organisational and Social Use. In N. Bessis & F. Xhafa (Eds.), *Next Generation Data Technologies for Collective Computational Intelligence* (Vol. 572, pp. Springer). Berlin: Springer.

Efron, M. (2011). Information search and retrieval in microblogs. *Journal of the American Society for Information Science and Technology, 62*(6), 996-1008. doi: 10.1002/asi.21512

Eisenhardt, K. M. (1989). Building Theories From Case Study Research. *Academy of Management. The Academy of Management Review, 14*(4), 532-550.

Engeström, Y. (1987). *Learning by Expanding: An Activity-Theoretical Approach to Developmental Research*. Helsinki: Orienta-Konsultit.

Engeström, Y. (2000). Activity Theory as a Framework for Analyzing and Redesigning Work. *Ergonomics, 43*(7), 960-974.

Fayyad, U. M., & Piatetsky-Shapiro, G. (1996). Advances in Knowledge Discovery and Data Mining. In U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth & R. Uthurusamy (Eds.), *Data Mining and Knowledge Discovery*.

Foot, K. (2002). Pursuing an Evolving Object: A Case Study in Object Formation and Identification. *Mind, Culture and Activity, 9*(2), 132-149.

Gao, H., Barbier, G., & Goolsby., R. (2011). Harnessing the Crowdsourcing Power of Social Media for Disaster Relief. *IEEE Intelligent Systems, 26*(3), 10-14.

Gayo-Avello, D. (2012). I Wanted to Predict Elections with Twitter and all I got was this Lousy Paper" -- A Balanced Survey on Election Prediction using Twitter Data. *arXiv:1204.6441v1*.

Goldberg, C. B. (2003). Applicant reactions to the employment interview: A look at demographic similarity and social identity theory. [doi: DOI: 10.1016/S0148-2963(01)00267-3]. *Journal of Business Research, 56*(8), 561-571.

Greenberg, J., & Méndez, E. (2009). *Knitting the Semantic Web*. London: Routledge.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199-220. doi: 10.1006/knac.1993.1008

Gruninger, M., & Fox, M. S. (1994). *The Role of Competency Questions in Enterprise Engineering.* Proceedings of the IFIP WG5.7 Workshop on Benchmarking - Theory and Practice, Trondheim, Norway.

Harris, M., & Clark JI, I. A. (2006). The Gene Ontology (GO) project in 2006. *Nucleic Acids Research, 34*, 322-326.

Hassan Ibrahim, N., & Allen, D. (2012). Information sharing and trust during major incidents: Findings from the oil industry. *Journal of the American Society for Information Science and Technology, 63*(10), 1916–1928. doi: 10.1002/asi.22676

Hine, C. (2000). *Virtual ethnography*. London: Sage.

Horrocks, I., & Sattler, U. (2005). *A tableaux decision procedure for SHOIQ.* Proceedings of the Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI), (pp. 448-453) Morgan.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons, 53*(1), 59-68. doi: 10.1016/j.bushor.2009.09.003

Kaptelinin, V. (2005). The Object of Activity: Making Sense of the Sense-Maker. *Mind, Culture, and Activity, 12*(1), 4-18.

Kim, H.-L., Breslin, J., Kim, H.-G., & Choi, J.-H. (2010). Social semantic cloud of tags: semantic model for folksonomies. *Knowledge Management Research & Practice, 8*, 193-202.

Kim, H. H. (2005). ONTOWEB: Implementing an ontology-based Web retrieval system. *Journal of the American Society for Information Science and Technology, 56*(11), 1167-1176. doi: 10.1002/asi.20220

Korpela, M., Mursu, A., & Soriyan, H. A. (2002). Information Systems Development as an Activity. *Computer Supported Cooperative Work, 11*, 111–128.

Kuhn, W. (2001). Ontologies in support of activities in geographical space. *International Journal of Geographical Information Science, 15*(7), 613-631. doi: 10.1080/13658810110061180

Langer, R., & Beckman, S. C. (2005). Sensitive research topics: netnography revisited. *Qualitative Market Research: An International Journal, 8*(2), 189-203.

Lenhart, A., Purcell, K., Smith, A., & Zickuhr, K. (2010). Social Media & Mobile Internet Use Among Teens and Young Adults. Washington: Pew Research Centre.

Leont'ev, A. N. (1978). *Activity, Consciousness, and Personality*: Prentice-Hall.

Liang, T.-P., & Turban, E. (2011). Introduction to the Special Issue: Social Commerce: A Research Framework for Social Commerce. *International Journal of Electronic Commerce, 16*(2), 5-13.

Lynch, C. F., Ashley, K. D., Aleven, V., & Pinkwart, N. (2006). *Defining "Ill-Defined Domains" A literature survey.* In V. Aleven, K. Ashley, C. Lynch & N. Pinkwart (Eds.). Proceedings of the Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains at the 8th International Conference on Intelligent Tutoring Systems, (pp. 1-10) Jhongli, Taiwan.

Millar, R., & Tracy, A. (1997). The employment interview. In O. Hargie (Ed.), *The handbook of communication skills* (2nd ed., pp. 453-479). London ; New York: Routlege.

Minsky, M. (1995). Steps to artificial intelligence. In G. F. Luger (Ed.), *Computation and Intelligence, Collected Readings.* (pp. 47–90). Menlo Park CA: MIT Press.

Nasukawa, T., & Yi, J. (2003). *Sentiment analysis: capturing favorability using natural language.* Proceedings of the 2nd International Conference on Knowledge Capture, (pp. 70 - 77).

Nelson, M. R., & Otnes, C. C. (2005). Exploring cross-cultural ambivalence: a netnography of intercultural wedding message boards. *Journal of Business Research, 58*(1), 89-95. doi: 10.1016/s0148-2963(02)00477-0

Noh, T.-G., Park, S.-B., Park, S.-Y., & Lee, S.-J. (2010). Learning the emergent knowledge from annotated blog postings. *Web Semantics Science Services and Agents on the World Wide Web, 8*(4), 329-339.

O'Leary, D. E. (2010). Enterprise ontologies: Review and an activity theory approach. *International Journal of Accounting Information Systems, 11*(4), 336-352. doi: DOI: 10.1016/j.accinf.2010.09.006

O'Leary, D. E. (2011). Blog mining-review and extensions: From each according to his opinion. *Decision Support Systems, 51*(4), 821-830.

Page, R. E. (2012). *Stories and social media : identities and interaction*. London: Routledge.

Parent, C., & Spaccapietra, S. (2009). Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization *Modular Ontologies* (pp. 5-23). Berlin: Springer-Verlag.

Park, H. W., & Thelwall, M. (2008). Link analysis: hyperlink patterns and social structure on politicians' web sites in South Korea. *Quality and Quantity, 42*(5), 687–697.

Qualman, E. (2009). *Socialnomics : how social media transforms the way we live and do business*. Hoboken, N.J: Wiley.

Raffler-Engel, W. v. (1983). *The Perception of Nonverbal Behaviour in the Career Interview*. Amsterdam: John Benjamins Publishing Company.

Rector, A. L., Zanstra, P. E., Solomon, W. D., Rogers, J. E., Baud, R., Ceusters, W., . . . Wagner, J. (1998). Reconciling users' needs and formal requirements: issues in developing a reusable ontology for medicine. *IEEE Transactions on Information Technology in Biomedicine, 2*(4), 229-242.

Rheingold, H. (2000). *The virtual community: homesteading on the electronic frontier*: MIT Press.

Roth, W.-M. (2007). Emotion at Work: A Contribution to Third-Generation Cultural-Historical Activity Theory. *Mind, Culture, and Activity, 14*(1), 40 - 63.

Rowe, M., Alani, H., Angeletou, S., & Fernandez, M. (2012). Community Analysis through Semantic Rules and Role Composition Derivation. *Journal Of Web Semantics*.

Schneider, S. M., & Foot, K. A. (2005). Web Shere Analysis: An Approach to Studying Online Action. In C. Hine (Ed.), *Virtual Methods: Issues in Social Research on the Internet*. Oxford: Berg.

Stock, W. G., & Weber, D. (2006). Facets of Informetrics. *Journal of Information, 57*(8), 385-389.

Thelwall, M. (2006). Interpreting social science link analysis research: A theoretical framework. *Journal of the American Society for Information Science and Technology, 57*(1), 60-68.

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology, 62*(2), 406-418. doi: 10.1002/asi.21462

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). Sentiment strength detection for the social web. *Journal of the American Society for Information Science and Technology, 63*(1), 163-173. doi: 10.1002/asi.21662

Thelwall, M., Sud, P., & Vis, F. (2012). Commenting on YouTube videos: From guatemalan rock to El Big Bang. *Journal of the American Society for Information Science and Technology, 63*(3), 616-629.

Thelwall, M., & Vaughan, L. B., L. (2005). Webometrics. *Journal of the American Society for Information Science and Technology, 39*(1), 81-135.

Tsui, E., Wang, W. M., Cheung, C. F., & Lau, A. S. M. (2010). A concept–relationship acquisition and inference approach for hierarchical taxonomy construction from tags. *Information Processing and Management, 46*(1), 44-57.

Turkle, S. (2005). *Life on the Screen: Identity in the Age of the Internet*: Simon & Schuster.

Uschold, M., & Grüninger, M. (1996). Ontologies: Principles, Methods, and Applications. *Knowledge Engineering Review, 1*, 96-137.

Wells, G. (2002). The Role of Dialogue in Activity Theory *Mind, Culture, and Activity, 9*(1), 43-46.

West, R., & Turner, L. H. (2011). *Understanding Interpersonal Communication: Making Choices in Changing Times* (2nd ed.). Boston: Cengage.

Yi, K. (2010). A Semantic Similarity Approach to Predicting Library of Congress Subject Headings for Social Tags. *Journal of the American Society for Information Science and Technology, 61*(8), 1658-1672.

Yoon, K. (2012). Conceptual syntagmatic associations in user tagging. *Journal of the American Society for Information Science and Technology, 63*(5), 923-935.

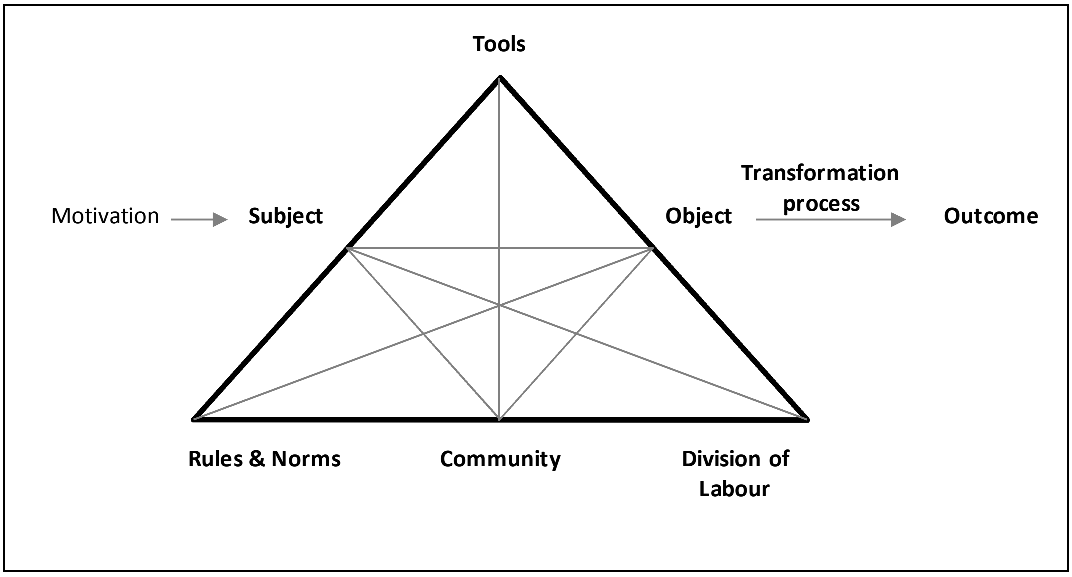

Figure 1: Activity Theory Structure (Engeström, 1987).

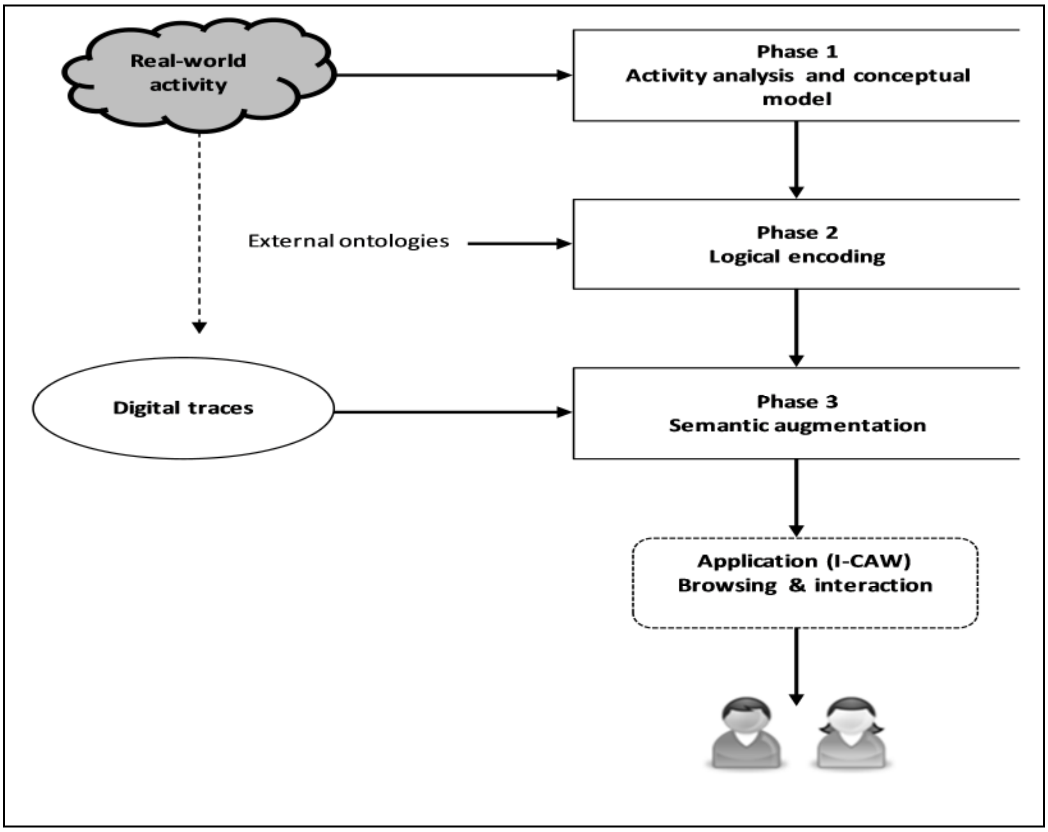

Figure 2: Overview of Methodology.

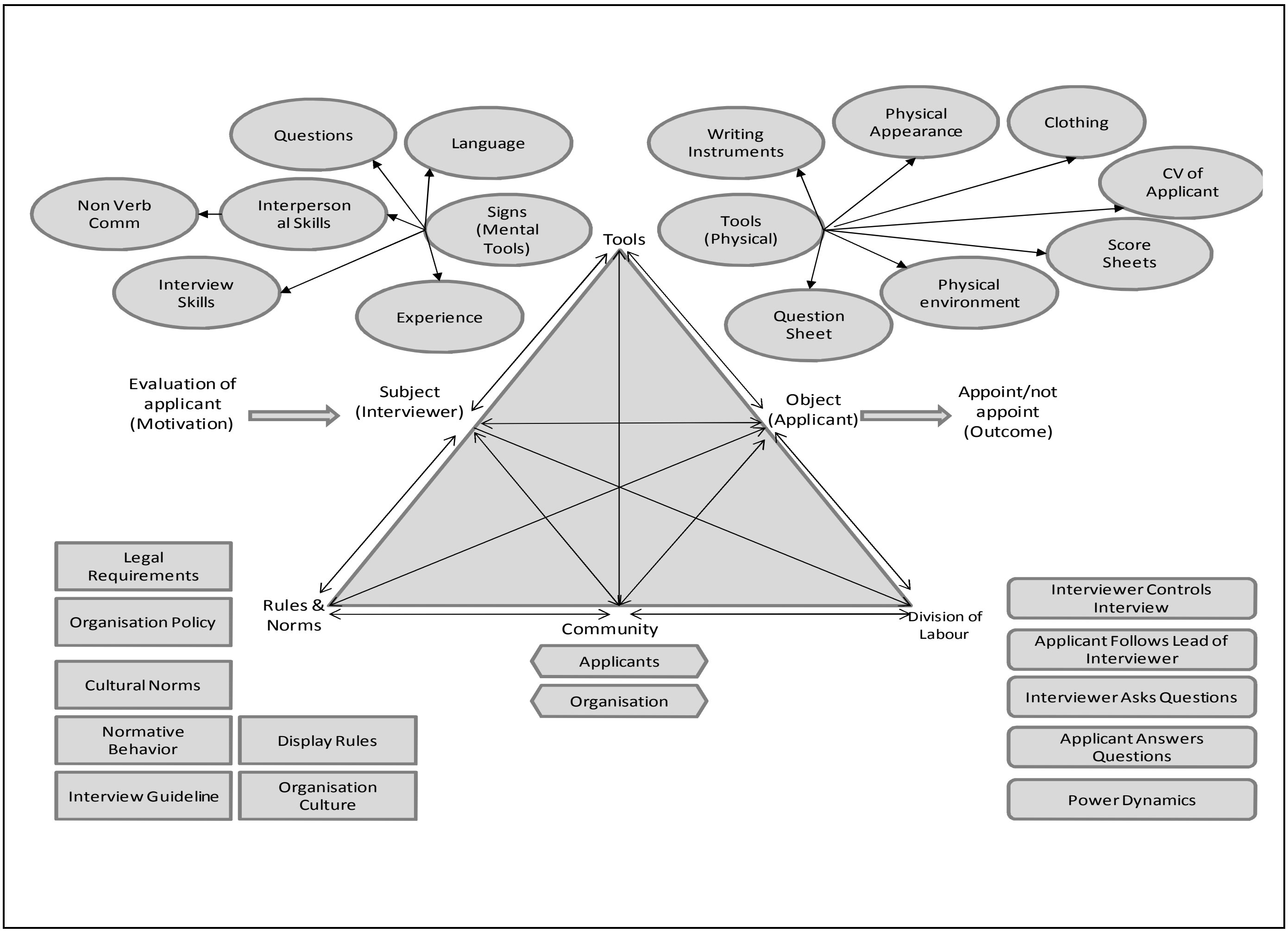

Figure 3: Conceptual Activity Model of the Job Interview.

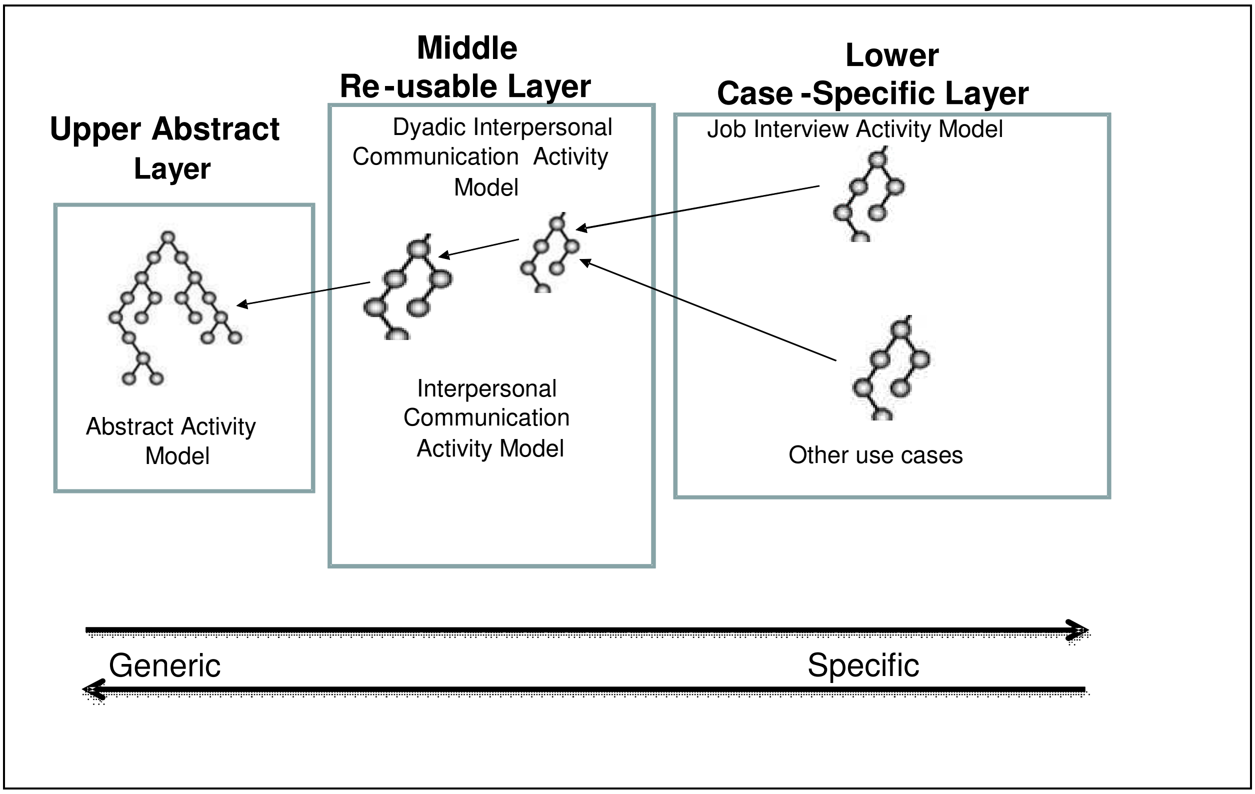

Figure 4: Multi-Layered Activity Model Ontology (AMOn).

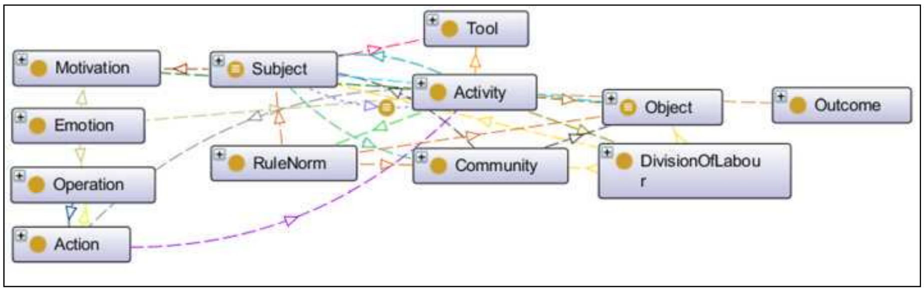

Figure 5: Upper Abstract Layer in AMOn (Modelling Activity System – cf. Figure 4).

Table 1: Upper Abstract Layer Concepts and Properties.

Controlled Natural Language (CNL) representation	**Description Logic (DL) representation**
Every Activity has Motivation.	Activity ⊆ ∃ hasMotivation.Motivation*
Every Activity has Tool.	Activity ⊆ ∃ hasTool.Tool
Every Activity has at least one Subject.	Activity ≥ 1 hasSubject.Subject
Every Activity is defined by an Activity Viewpoint that is viewpoint of Subject.	Activity ⊆ ∃ isDefinedBy.ActivityViewPoint ∩ ∃ isViewPointOf.Subject
hasMotivation is a relationship.	∃ hasMotivation ⊆ Activity T ⊆ ≡ has Motivation. Motivation
has Action is a relationship	∃ has Action ⊆ Activity T ⊆ ≡ has Action.Action has Action ≡ isPartOf.Activity

* ⊆: Concept inclusion (i.e. subclass of); ∃: existential restriction (i.e. there exists); ≥: at least one; ≡: equivalence, i.e. is equal to; T: thing; ∩: intersects with.

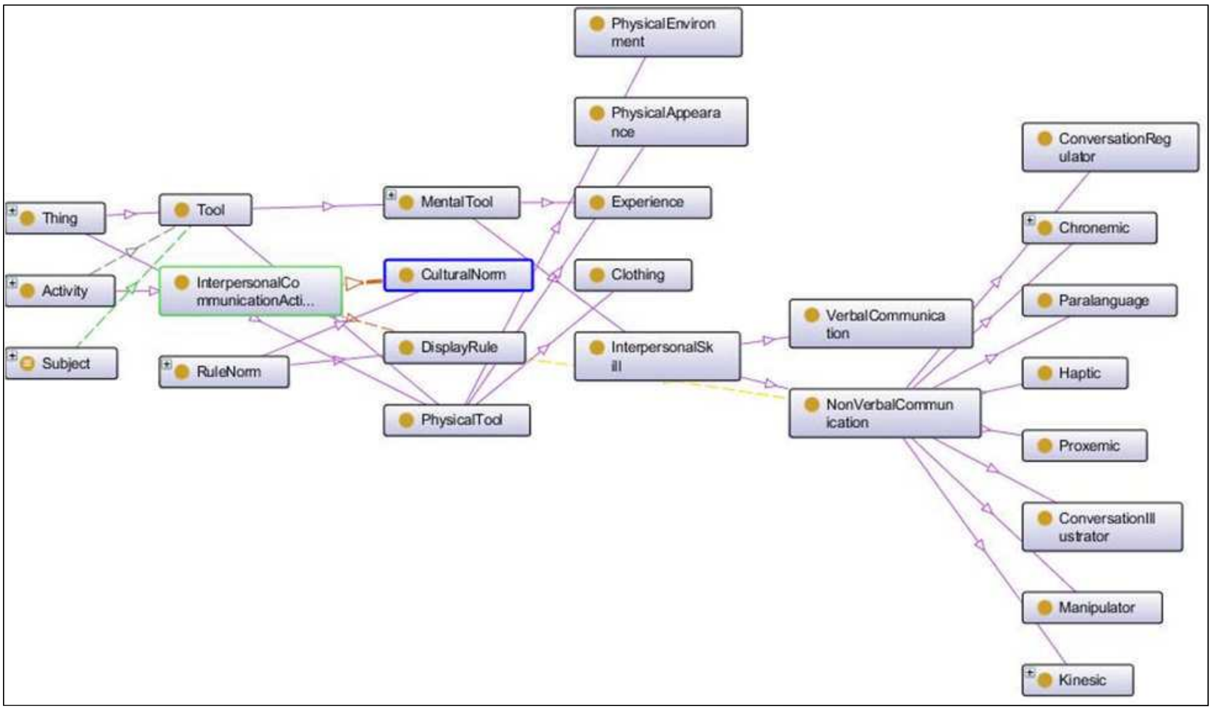

Figure 6: Interpersonal Communication Ontology Module (Interpersonal Communication Related Concepts from the Conceptual Model in Figure 3).

Table 2: Middle Layer Concepts and Properties.

Controlled Natural Language (CNL) description	Description Logic (DL) description
Interpersonal Communication Activity is part of Activity	InterPersonalCommunicationActivity ⊆ Activity
Every Mental Tool is kind of Tool.	MentalTool ⊆ Tool
Every Interpersonal Skill is a kind of Mental Tool.	InterpersonalSkill ⊆ MentalTool
Non verbal Communication Skills (NVC) is part of Mental Tools	NVC ⊆ Mental Tool
Body Language is part of NVC	BodyLanguage ⊆ NVC

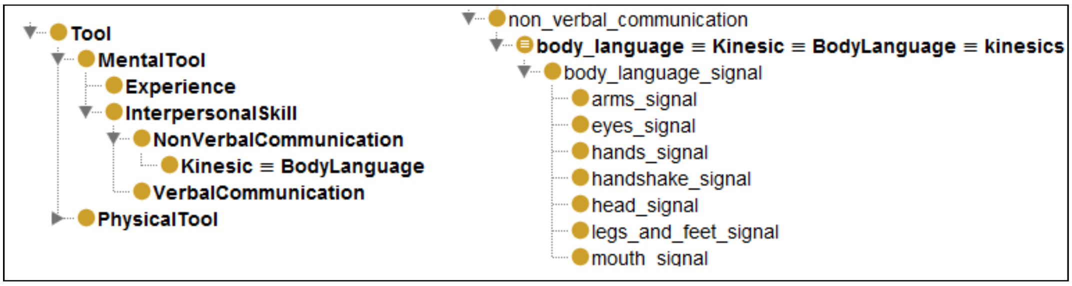

Figure 7: Utilizing External Ontologies to Specialise Body Language Concepts in the Middle layer. Left: Before Expansion (as in Conceptual Model - Figure 4), Right: After Expansion.

Table 3: AMOn Features.

Feature	**Value**
Number of classes	190
Number of properties	60
Number of Individuals	1043
DL Expressivity	SHOIQ
Number of Axioms	4431

Table 4: Coverage of Concepts by AMOn* (Total 143 Keywords).

No. of ***exact*** keywords covered by concepts from activity model only	26
No. of ***exact*** keywords further covered by adding external ontologies	4
No. of ***synonymous*** keywords covered by concepts from activity model only	34
No. of ***synonymous*** keywords further covered by adding external ontologies	6
No. of keywords covered by ***abstract*** concepts from activity model	96
No. of keywords further covered by ***abstract*** concepts after adding external ontologies	7

* Exact/synonymous matches were also covered by matches at abstract level

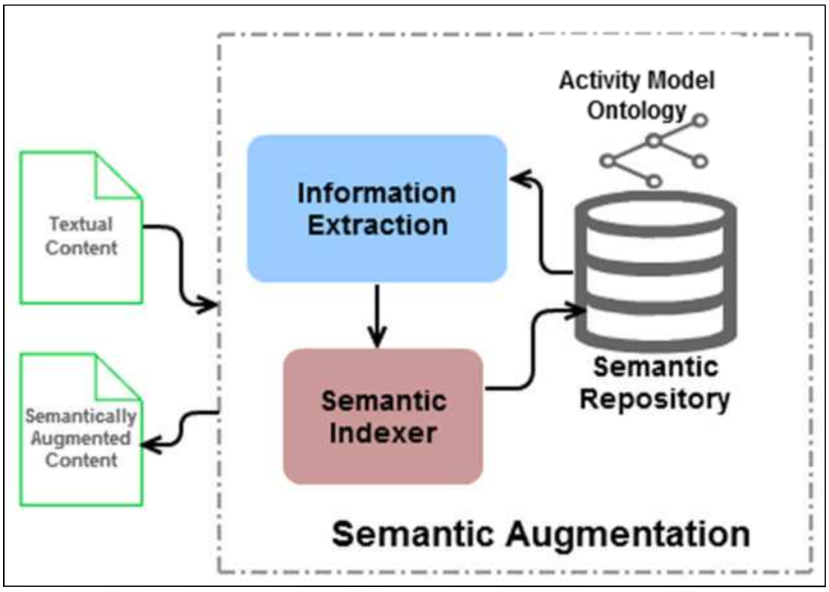

Figure 8: Semantic Augmentation Service Architecture.

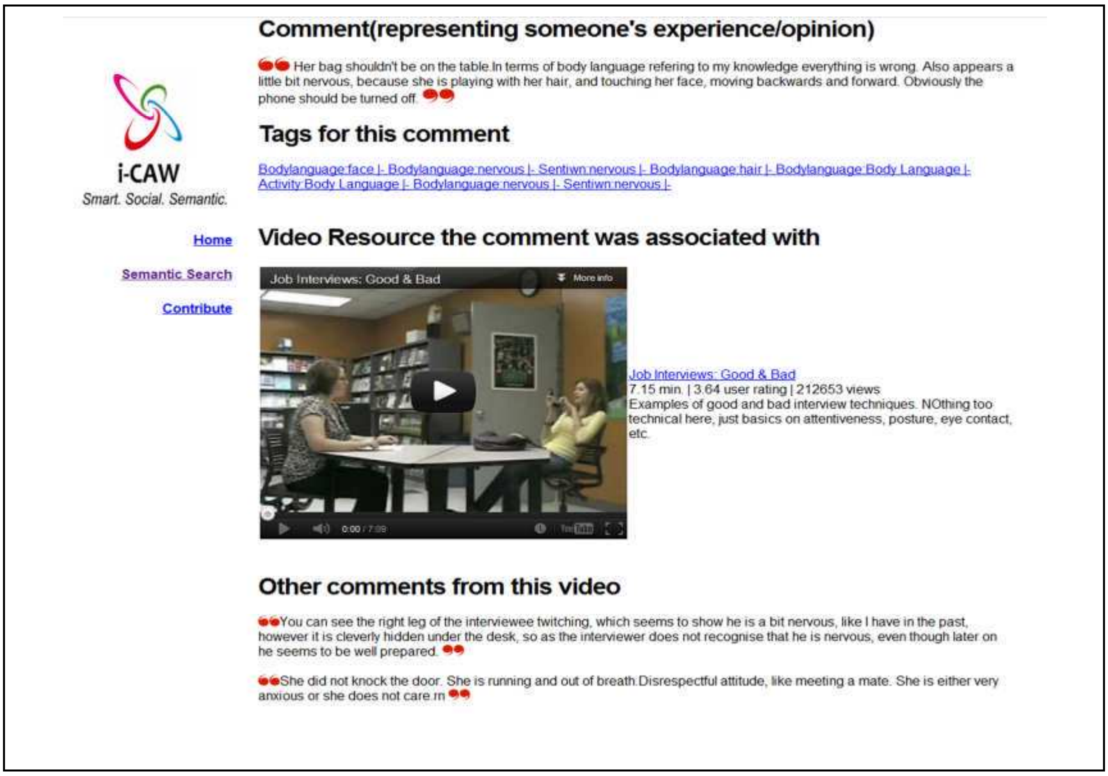

Figure 9: Screen Shot Showing Content Presented to User.

[i] http://imash.leeds.ac.uk/imreal/wp3/amon.html

[ii] ROO is underpinned by an ontology methodology, and provides a 'wizard-like' interface for defining ontology constructs in a controlled natural language.

[iii] http://imash.leeds.ac.uk/ontologies/amon/BodyLanguage.owl

[iv] http://imash.leeds.ac.uk/ontologies/amon/SentiWordNet.owl

[v] http://imash.leeds.ac.uk/ontologies/amon/WNAffect.owl

[vi] http://xmlns.com/foaf/spec/

[vii] The AMOn ontologies are published for reuse http://imash.leeds.ac.uk/ontologies/amon/

[viii] http://www.ontotext.com/owlim

[ix] http://gate.ac.uk/

[x] http://www.openrdf.org/doc/sesame2/api/

[xi] This service is published for reuse http://imash.leeds.ac.uk/services/semanticservices/augmentation.html

[xii] http://imash.leeds.ac.uk/services/i-caw/