A Computational Framework for Identity and Its Web-based Realization

William Nick
Madison Mccotter
Siobahn Day
Marguerite McDaniel
North Carolina A&T State U.
Greensboro, NC 27411
{wmnick, mumccott, scday, mamcdan2}
@aggies.ncat.edu

Emma Sloan
Brown University
Greensboro, NC 27411
emma_sloan2@brown.edu

Hannah Foster
James Mayes
North Carolina A&T State U.
Greensboro, NC 27411
hnfoster@aggies.ncat.edu
jpmayes@ncat.edu

Albert Esterline
North Carolina A&T State U.
Greensboro, NC 27411
esterlin@ncat.edu

Abstract

This paper presents a computational framework for identity and is particularly focused on identifying the culprit in a crime-scene investigation. A case is conceptualized as a constellation of situations in the sense of Barwise's situation theory. Data on a case is stored as RDF triples in a triple store. Several relevant OWL ontologies have been developed and supplemented with SWRL rules. Uncertainty and combining levels of (possibly conflicting) evidence are handled with Dempster-Shafer theory. A webpage is being developed to make available to students of criminal justice the results of our work. The user will be able to query about evidence and follow how it accrues to various hypotheses.

1 Introduction

We are developing a computational framework for identity, and, with our focus on criminal investigations, we are integrating ontologies and (to combine levels of evidence) Dempster-Shafer theory into our system. A particularly important goal of this project is a web interface to this system for learning purposes. This webpage will allow a student to query information regarding criminal investigations and follow how evidence accrues to various hypotheses about the identity of the culprit.

The SuperIdentity project is the state-of-the-art in frameworks for identity (Creese et al.). It starts with some known information or *element* of identity, such as a username or email address, and *transforms* that element into others, e.g., by looking up an email address to find the associated username. These elements are grouped by type (e.g., phone number) into *characteristics*, multisets of elements. The set of all characteristics is a person's *superidentity*, a compilation of all known information on them. Our framework covers all aspects of the SuperIdentity framework but from a situation-theory perspective. We assemble constellations of situations (in the technical sense of Barwise and Perry (Barwise and Perry 1983)) to produce a case as in the legal sense, providing more structure and provenance than provided by superidentities.

The remainder of this paper is organized as follows. Section 2 presents background: situation theory (the theoretical background for our representations), Semantic Web resources (our OWL ontologies serve as knowledge bases, and data is stored as RDF triples conforming to our ontologies), and Dempster-Shafer theory (used to handle uncertainty and collaborating and conflicting evidence). Section 3 presents our running example, Section 4 presents our ontologies, and Section 5 summarizes the encoding of our examples in RDF. Section 6 discusses the SWRL (Semantic Web Rule Language) rules that complement our ontologies. Section 7 addresses evidence in the legal sense and the importance of certain objects, viz., biometric artifacts, that persist across situations. Section 8 presents our application of Dempster-Shafer theory, Section 9 outlines a functional design of our webpage, and Section 10 sketches the ongoing implementation of our web-based system. Section 11 concludes and suggests future work.

Our framework is used in a way compatible with contemporary crime-scene investigation, with most information manually encoded as RDF triples and possibly automated encoding of documents. We rely on human perception except for biometric matching. Ontologies, which capture expert knowledge and conventional practice (there is no machine learning), constrain the encoding and support inference. Dempster-Shafer theory reveals how evidence combines and provides guidance even when evidence is weak. For other recent presentations of our framework, see (McDaniel et al. 2017a) and (McDaniel et al. 2017b) regarding ontologies and (Sloan et al. 2016) and (Sloan et al. 2017) regarding application of Dempster-Shafer theory.

2 Background

2.1 Situation Theory

Our computational framework is based on the situation theory of Barwise and Perry (Barwise and Perry 1983), especially as systematized by Devlin (Devlin 1995). According to Barwise, "'[s]ituation' is our name for those portions of reality that agents find themselves in, and about which they exchange information" (Barwise 1989). A situation supports elementary items of information, called *infons*, each essentially a relation among objects at a time and place (or possibly the lack of such a relation). A *real situation* supports an indefinite number of infons. We generally work with *abstract situations*, each supporting a finite number of (possibly parameterized) infons. An abstract situation amounts to a type under which real situations are classified.

There are constraints between situations, as expressed, for example, by "smoke means fire." By virtue of constraints, one situation may carry information about another. By virtue of conventional (linguistic) constraints, an *utterance situation*, in which someone performs a (declarative) speech act, carries information about a *described situation*. Whether the speech act is felicitous may depend on *resource situations* related by conventions to the utterance situation; in a purely linguistic setting, such situations typically support infons expressed by relative clauses.

In our framework, an *id-action* takes place in what we call an *id-situation*. Any id-action is considered an assertion of identity even if it is not verbal. So an id-situation is an utterance situation, and the crime scene is the corresponding described situation. Supporting situations essential to crime-scene evidence (e.g., those where suspects' fingerprints were recorded) are resource situations: there are conventional constraints requiring the existence of properly executed situations for the evidence to be admissible. Together, the id-situation, the described situation (crime scene), and the resource situations make an *id-case*.

2.2 Semantic Web Resources

The semantic web is built off of two W3C standards (Pan 2009): the resource description framework (RDF) [ref] and the resource description framework schema (RDFS). The web ontology language (OWL) extends the expressiveness of RDFS and allows for the creation of ontologies. "Ontology" is a term borrowed from philosophy, where it means the conceptualization of entities in the world and how they interact with each other, but, in computer science, it denotes a conceptualization of a domain.

RDF is a W3C recommendation that allows for the annotation of web resources. RDF statements (known as triples) are in the form of *subject predicate object*, where *predicate* is a binary relation. A resource (thing) is denoted in RDF by a uniform resource identifiers (URI), a string unique across the web. A URI reference (URIref) is a URI with an optional fragment identifier at the end. A URIref is usually represented as a Qname, `pre:lp`, where `pre` is a URI (essentially a namespace pefix) and `lp` is the local part. A blank node (*bnode*) is a resource that is not identified by a URIref, functioning like a pronoun. One RDF serialization defined by the W3C is N3, in which triples are expressed by the three components separated by whitespace. If several triples share a subject, we can abbreviate by listing the common subject then listing predicate-object pairs separated by semi-colons.

RDFS allows for the definition of new RDF classes and properties. An individual is an instance of a class. A class may be a subclass of classes and a property may be a subproperty of properties. A property has a *domain*, which is a class to which its subjects belong, and a *range*, which is a class or datatype to which its objects belong. For a succinct representation, where *`p`* is a property, *`Dom`* is its domain, and *`Rng`* is its range, we write *`p: Dom`* $\rightarrow$ *`Rng`*. *Object properties* have classes as ranges while *datatype properties* have datatypes. Unlike RDFS, OWL allows for the expression of local scope of properties, disjointedness of classes, Boolean combinations of classes, cardinality restrictions, and special characteristics of properties.

SPARQL (Pérez, Arenas, and Gutierrez 2006) is a SQL-like query language for triple stores. The `WHERE` clause contains a pattern of triples that will be matched by the query engine. Query output is what is bound to the variables in the `SELECT` clause. Various applications allow one to infer new triples from those present in the triple store via connections captured in the OWL ontologies. For additional inference patterns, the ontologies can be supplemented with rules in the Semantic Web Rule Language (SWRL). These are if-then rules that use the concepts expressed in the ontologies.

2.3 Dempster-Shafer Theory

Dempster-Shafer (DS) theory is a justification-based way of distributing evidence (Halpern 2005). "Mass" of evidence distributes to sets of elements or outcomes, with unassigned mass, representing ignorance, given to the set of all elements, called the *frame of discernment*. This assigns masses sum to 1.0. A *focal element* is a set with non-zero mass. A *refinement* is the analysis on the frame of discernment to get a more detailed frame of discernment.

Given a mass function, the *belief* function for a set is the lower bound for likelihood, calculated by adding the masses of all subsets of the set, while the *plausibility* is the upper bound for the likelihood, calculated by adding the masses from all of the sets that overlap the set. In symbols, for a frame of discernment Θ, a mass function m, and any subset θ of Θ, $Bel(\theta) = \sum_{\theta^* \subseteq \theta} m(\theta^*)$ and $Plaus(\theta) = \sum_{\theta^* \cap \theta \neq \emptyset} m(\theta^*)$. Thus, for any $\theta \subseteq \Theta$, $Bel(\theta) \leq Plaus(\theta)$.

DS theory allows for the combination of multiple mass functions for different kinds of evidence to produce a new mass function that relates to the combined evidence. There are a number of combination rules that fit different types of data better. For example, Dempster's rule combines pieces of evidence that are equally reliable while preserving the uncertainty inherent within each piece of evidence. Specifically, Dempster's rule calculates a measure, *K*, of conflict between the mass functions and divides that measure as mass among the different focal elements, including the focal element that is the entire frame of discernment. In symbols, with $K = \sum_{B \cap C \neq \emptyset} m_1(B)m_2(C)$ for mass functions m_1 and m_2 and focal elements B and C. Dempster's rule combines m_1 and m_2 as $m_{12} = (\sum_{B \cap C = A} m_1(B)m_2(C))/(1-K)$. Other combination rules include Zhang's rule (which allows combination of mass functions with different frames of discernment) and the mixing rule (which assigns different weights to the combined mass function and so compensates for the fact that some pieces of evidence may be more reliable than others). And often researchers create their own, context-specific combination rules.

3 Running Example

We consider a scenario where a theft has occurred at a party. There is a list of possible suspects in the form of a guest list. Evidence from the crime scene includes a group photograph from a security camera with one guest with their hand on the doorknob of the door to where the valuables were kept and a fingerprint on that doorknob. This scenario is a constellation of situations, centering around two separate id-situations for the two pieces of evidence: the fingerprint and the snapshot. Situation s_1 (see Figure 1) is the id-situation for the fingerprint case; an analyst compares fingerprints on file from the partygoers to the forensic one at the crime scene. In situations s_{3a}-s_{3d}, a suspect has their fingerprint taken by a police officer, and in situation s_4, the criminal touches the doorknob, placing the forensic fingerprint. The CSI team lifts the fingerprint from the doorknob in situation s_5. The id-situation for the case with the security camera image, s_2, is supported by its own constellation of situations. See Figure 2. Police take a mugshot of each suspect in situations s_{6a}-s_{6d}. Those mugshots are then compared with the security camera image in s_2. The security camera records the group in situation s_7, which acts as an utterance situation, describing the group in s_8. Situation s_4, in which the fingerprint is left by touching the doorknob, is a part of situation s_8.

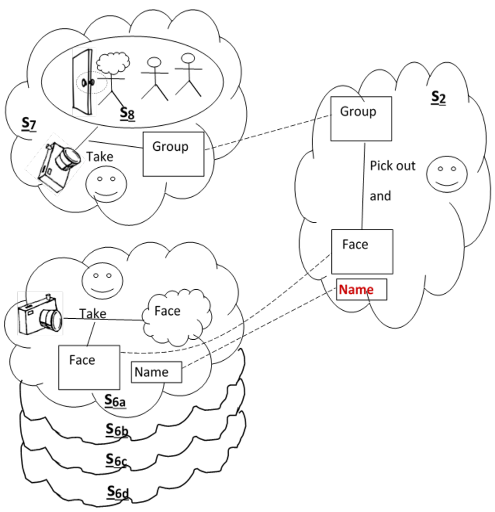

Figure 1: Mugshot ID case

4 The Ontologies

Figure 3 shows the ontologies created for the framework and their relationships to each other. Each ontology inherits from the ontology below.

Our ID-Situation Ontology (to which we associate the prefix `id`) focuses on situations and constellation of situations (i.e., id-cases) that involve id-actions as well as any evidence supporting them. This ontology is built on the Situation Ontology (to which we associate the prefix `sit`), whose two major classes are `sit:Situation` and `sit:Infon`, whose children are essential in encoding our example. Subclasses of `sit:Infon` corresponding to various relations are defined in the ID-Situation Ontology. For such a subclass, we define properties with it as domain for the argument positions in the corresponding relation. This avoids RDF's restriction to binary relations ("properties") and accommodates variable-arity relations. For simplicity, we associate time and location with a situation

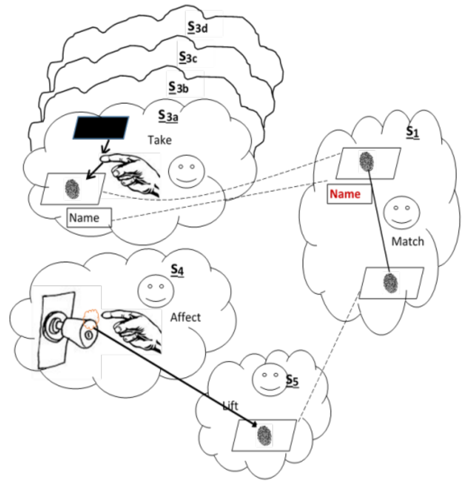

Figure 2: Fingerprint ID case

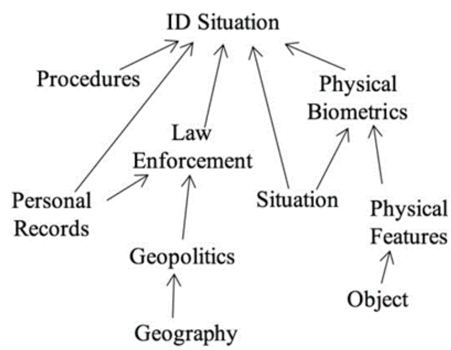

Figure 3: The ontologies and their inclusions

rather than with individual infons. Consequently, we define top-level classes `sit:Temp` and `sit:Loc` as well as various functional properties such as `sit:tempLoc: sit:Situation` $\rightarrow$ `sit:Temp` and `sit:spatialLoc: sit:Situation` $\rightarrow$ `sit:Loc`. For the `sit:Situation` class, there is a reflexive and transitive object property `sit:partOf: sit:Situation` $\rightarrow$ `sit:Situation` to indicate that one situation is a part of another.

The ID-Situation Ontology includes class `id:IdCase`, an instance of multiple situations that form an id-case. There is an object property `id:hasSituation: id:IdCase` $\rightarrow$ `sit:Situation`, connecting an id-case to its constituent situations. Subproperties of `id:hasSituation` are `id:hasIdSituation: id:IdCase` $\rightarrow$ `sit:Situation` and `id:hasSupportingSituation: id:IdCase` $\rightarrow$ `sit:Situation`. The first acts as a functional property that relates an id-case to its id-situation; the latter links an id-case to supporting situations. We also have an equivalence property, `id:coordindatedIdCase: id:IdCase` $\rightarrow$ `id:IdCase`, that relates id-cases that refer to the same scenario.

As shown in Figure 3, the ID-Situation Ontology incorporates other ontologies that relate not only to the structure of a case but also to the specific kind of information–biometric artifacts–and procedures needed for evidence to support id-actions. This includes the Physical Biometric Ontology that addresses biometric artifacts, which are images of the suspects' physical features registered for use by forensic professionals. For the information captured by physical biometrics, we have a Physical Features Ontology that addresses the human body, which relates specific surface features to specific persons, allowing the biometric images to serve as identifiers.

The most important of the supporting ontologies is the Law Enforcement Ontology. (For simplicity, we now omit prefixes.) The standard FOAF ontology has a top-level `Agent` class with children `Organization` and `Person`. We provide a child `LawEnforcementAgency` of `Organization` and a child `LawEnforcementProfessional` of `Person` itself with children `ForensicProfessional`, `PoliceOfficer`, `PoliceInvestigator`, and `ProsecutionProfessional`. There is an affiliation property associating agents with organizations and a certification property associating forensic professionals with certificates. There is a Personal Records Ontology.

5 Encoding in RDF of the Example

We outline the RDF encoding of the fingerprint case, providing an example of code. A shorter summary of the mugshot case is presented.

5.1 The Fingerprint Id-Case

s_1 is the id-situation for the fingerprints. It supports two essential infons, both instances of children of `id:MatchingFpInfon` (a child of `sit:Infon`). One child of this class is `id:AnalystMatchingFpInfon`, information that an attempt is made to match a fingerprint from the scene against a recorded fingerprint, and the other is `id:SimilarFpInfon`, information on the similarity measure for the match and the matching procedure used. For the first suspect, s_1 supports the following two infons (denoted by blank nodes, with "_" in place of a prefix).

```
_:i1la a id:AnalystMatchingFpInfon;
     id:fpAnalyst forprof:117;
     id:fpObserved forensicfp:652;
     id:fpRecorded fpfile:496;... .
_:i1a a id::SimilarFpInfon;
    id:fpObserved forensicfp:652;
    id:fpRecorded fpfile:496;
    id:simMeasure "0.92";
    id:simProc similar:Proc1; ... .
```

We assume that the relevant agency has indexed the individuals with numerical identifiers. We introduce prefixes for individuals thus indexed: `forprof` for forensic professionals, `forensicfp` for fingerprints collected at crime scenes, and `fpfile` for fingerprints on file. We also assume unique identifiers with prefix `similar` for the matching procedures used. The same pair of fingerprints appears in both infons. There are similar pairs of infons for the other suspects.

s_{3a} is the situation where the fingerprints of the first subject were taken and recorded. It supports one essential infon, the information that a forensic professional takes the fingerprint of a subject. There is one such infon for the other three fingerprint-recording situations. s_4 is the described situation, where someone leaves their fingerprint on the doorknob. There are two essential items of information here, that the fingerprint is on the doorknob, and that some suspect left their fingerprint. s_5, where the fingerprint on the doorknob is lifted, has one essential item of information, that a forensic professional lifts a fingerprint.

5.2 The Mugshot Id-Case

s_2, the id-situation for the mugshots, supports two essential infons similar to those in the fingerprint id-situation, s_1. The photo of the culprit is a part of the photo that the officer took: the part showing their face. s_{6a}, where the mugshot of the first subject was taken and recorded, supports one essential infon, that a given forensic professional takes and records a mugshot of a suspect with a given camera. There is one such infon for each of the remaining three situations. s_7, where a forensic professional takes a picture, has two significant infons. One is that a certain officer takes a picture of a given situation with a given camera thereby producing a given group photo. s_7 is an utterance situation: it produces an artifact carrying information about another situation, viz., s_8. The other significant infon in s_7 is that a certain group is in the described situation, s_8. When our culprit is identified, we add a triple stating that they are a member of the group. The described situation, s_8, has one significant infon, for the touching. Once we have identified the culprit, we add a triple for the toucher. This infon is supported by the described situation while s_7 carries this infon by virtue of the photo it produces. There is a part-whole relation between s_4 and s_8.

6 SWRL Rules

SWRL rules allow us to infer new triples and thus fill in our descriptions of objects, situations, and agents based on triples already in our ontologies. There are two significant tasks for our SWRL rules: identify the culprit and classify situations and entire id-cases.

6.1 Identifying the Culprit

Typically, there is a described situation where the culprit is unidentified and an id-situation, where the evidence is presented for pronouncing a judgment on the identity of the culprit in the described situation. For example, in the fingerprint case, once we have an identity judgment, we can fill in the value for the `id:fpProducer` property for the instance of `id:LeaveFpInfon` supported by s_4. We have created a rule that does the updates subsequent to identifying the culprit. This includes supplying `id:toucher` in `id:TouchInfon` and an agent in `id:AgentInfon` for the described situation s_8. We also assert a triple of the form `x sit:agentInSit s`, indicating that `x` is the agent of interest in situation `s`. This is top-level information not associated with any other situation that provides one way of identifying the agent. Inference will generally identify several suspects as "the" culprit since inference does not consider the goodness of biometric matches; we handle level of evidence with Dempster-Shafer theory – see Section 8.

6.2 Classifying Situations and Id-Cases

We need abstract situations as types to classify real situations and abstract id-cases to classify constellations of situations in a way conducive to identification. The ID-Situation Ontology has subclasses of class `sit:Situation`, essentially abstract situations, and it has an `id:IdCase` class, which has subclasses for classification. Determining whether a given situation should be an instance of a given situation class is a classification problem that hinges on whether the real situation supports certain infon subclasses. When we described our running example, we described real situations, but the descriptions themselves, where we talked about essential infons, basically formulated abstract situations. Our classifying SWRL rules, then, have the form

```
Situation(?s), ... -> SituationSubClass(?s)
```

The conditions that fill in the ellipsis relate to the infons that `?s` supports. We also classify an instance of `id:IdCase` as an instance of a subclass of that class. Finally, we have a SWRL rule for determining that an instance of the mugshot id-case and an instance of the fingerprint id-case are coordinated

7 Biometric Artifacts and Legal Evidence

Objects and situations (or events) are complementary (Galton and Mizoguchi 2009). Objects are created, changed, copied, and destroyed in situations, and situations consist of objects related in various ways. Though our foundations are built on situations, objects are important in several areas. Some objects are passive participants (e.g. doorknob) while others play essential roles in capturing evidence (e.g. camera). To support our conclusions from the evidence collected, we may need facts regarding these objects. The objects of primary interest are biometric artifacts, serving as threads that stitch together the situations building an id-case. E.g., in the fingerprint id-case, fingerprints are recorded on file (situations s_{3a}-s_{3d}), providing objects used in the id-situation, s_1, and a fingerprint is lifted from the doorknob in s_5, providing the object against which the fingerprints on file are compared.

To see how such objects count as evidence, note that evidence in criminal prosecution includes any documents, testimony or tangible objects that tend to prove or disprove the existence of alleged facts (Black, Garner, and McDaniel 1999). Documentary evidence consists of any written object or article, e.g., letters, contracts, deeds, licenses, and certificates, presented as proof. Testimonial evidence consists of any statement made under oath by a witness during trial or at deposition. "Tangible object" evidence refers to any physical item or its representation presented as proof to support an alleged fact. Physical evidence includes biological and non-biological trace evidence. Here, trace evidence is defined as evidence that can be transferred between people, objects or the environment. Physical evidence also includes facial recognition (photography and videography) and fingerprint and biometric (DNA, blood, semen, saliva, urine, feces, hair, teeth, bone, tissue, and cells) evidence. The use of all evidence is subject to law, legal rules and procedures to determine its admissibility and probative value. Additionally, the provenance, preservation (cf. chain of custody–CoC–immediately below), and analysis of evidence is central to its forensic application.

In particular, biometric artifacts used as evidence have to be genuine throughout. CoC theory addresses what is essential "to ensure the integrity of evidence" (Giannelli 1993). Our framework facilitates application of CoC theory, focusing (for now) on physical, or "real," evidence, tangible evidence used to prove a fact that is at issue in a case (Citizens Information 2014). It also has to be relevant, material, and competent (Findlaw 2016) to be admitted in court. Following what CoC theory states is required, one can authenticate real evidence since CoC theory requires the mapping of who, what, where, and how evidence is obtained and handled (Giannelli 1993). We appeal to CoC theory since our central focus is to evaluate whether the metadata or forensic data (as real evidence) is sufficient to identify suspects. Our framework records CoC steps followed; it does not itself physically obtain or preserve the evidence.

8 Application of Dempster-Shafer Theory

For each id-case, a numerical measure is created of who is likely to be the criminal and then all known id-cases are combined using Dempster-Shafer theory. For each id-situation and each suspect, there is a distance measure that is part of the id-situation. A mass function is then created based on that distance measure. To work with the scenario from the running example, to create the mass function for the fingerprint id-case, as shown in Figure 1, masses are assigned based on the distance between each suspect's fingerprint and the fingerprint from the crime scene. Conversion of distances to masses uses a customized sigmoid function, which provides a threshold below which a possible match can be ignored. Those masses are then normalized so that all values sum to 1.0.

An id-situation is tied to resource situations through constraints. Every piece of evidence in the id-situation has been collected in some other situation or set of situations, following appropriate legal procedures to maintain a chain of custody, a convention that specifies a related situation or set of situations (e.g., the s_3 situations in Figure 1) that must have occurred. There are three possible interpretations of the constraints in the context of Dempster-Shafer theory. One is that each set of situations provides a separate mass function so that the mass function from the resource situations must be combined with the mass function from the id-situation. The second is to consider each resource situation as a refinement of the frame of discernment created in the id-situation. The third interpretation has the resource situations modifying the mass function.

Treating each set of situations as a separate piece of evidence, with its own mass function, would sanction application of Dempster-Shafer combination rules. The frame of discernment for the id-situation in fact is the same as the frame of discernment for a collection of resource situations that covers all the suspects from the id-situation. Dempster's combination rule is appropriate because the underlying uncertainties should be preserved.

Treating the constraints as refinements would make each group of resource situations modify the arrangement of the set of suspects. As noted, however, in our framework, the frame of discernment is the same throughout.

Not all resource situations can be accommodated using the two previously described methods. For example, in Figure 1, information from situation s_5 modifies the id-situation, but it does not refer to any suspect and so does not refine the frame of discernment or even rearrange focal elements. The analysts who collected the crime scene fingerprint might be untrustworthy. This could be handled by moving some mass to the entire frame of discernment or by weighting the piece of evidence less heavily when combining it with evidence from other id-situations.

Table 1 shows the mass, belief, and plausibility for the mass functions from the two id-situations and their combination. Masses of non-singleton sets other than the entire frame of discernment ("All") are all zero. The mass values have been subject to modification to accommodate aspects of the resource situations as just discussed.

Photographic Evidence			
Suspect	Mass	Belief	Plaus
201	0.399	0.399	0.595
202	0.405	0.405	0.601
203	0	0	0.196
204	0	0	0.196
All	0.196	1.0	1.0
Fingerprint Evidence			
Suspect	Mass	Belief	Plaus
201	0.290	0.290	0.596
202	0	0	0.306
203	0.215	0.215	0.521
204	0.188	0.188	0.494
All	0.306	1.0	1.0
Combined Evidence			
Suspect	Mass	Belief	Plaus
201	0.528	0.528	0.636
202	0.222	0.222	0.330
203	0.076	0.076	0.183
204	0.066	0.066	0.174
All	0.108	1.0	1.0

Table 1: Mass, belief, and plausibility (abbreviated to plaus) measures for the three mass functions created by our two id-situations and their combination using Dempster's rule

9 Webpage

This section outlines the functionality of the webpage interface to the system we are building. The webpage will have three major functional areas: the case description , query, and evidence panes.

9.1 Case Description Pane

The case description pane will include a menu of cases. When a case is clicked, a short paragraph describing the case will appear, providing information about the case including an outline of the events as well as a summary of the kinds of evidence available. The user may switch between the various scenarios and investigate the scenario of their choice. The case description will have a button to bring up a menu of suspects. When the user selects a given suspect, a short description of the suspect appears along with a template for constructing queries about the suspect. Generally, this pane provides information to initiate interaction in the other two panes.

9.2 Query Pane

This pane, whose content is specific to the selected case, supports both queries on the triple stores and inferences made on these stores given the ontologies and the SWRL rules. There is a tab for each kind of evidence. In our running example, there is a tab for the fingerprint evidence and a tab for the mugshot evidence. There is also a tab for inferring the identity of the culprit, and a tab for classifying the case and its constituent situations.

The contents of the tabs for the various kinds of evidence are structured similarly. Consider, for example, the tab for the fingerprint evidence. Often a value will be selected from a menu. The following are some possible topics for queries.

- Who took a given fingerprint that is on file? When? Where?
- How was the fingerprint preserved or copied?
- Who lifted the fingerprint?
- Who handled the comparison of the fingerprint on file and the forensic fingerprint?
- How good a match is the match between a given fingerprint on file and forensic fingerprint?

Many of these queries relate to provenance, including chain of custody. Note that it is natural for a query to span situations, often by following the CoC. For example, one could ask for the name of the professional who took the fingerprint on file in situation s_{3a} that is used in the id-situation s_1. Many queries could be issued for a set of fingerprints on file. The results will be displayed in a table, possibly ranked by the value of some field (e.g., the goodness of match). If the result of a query indicates some suspect is particularly interesting, the user may bring up the template mentioned above for queries about a person. Legal professionals may also be of interest (e.g., those who took fingerprints), and a variation of the template will be available for querying about them.

The tab for inferring the identity of the culprit basically does just that. Often, however, there will be more than one culprit inferred since mere inference does not take into account how good the evidence is (which is the realm of Dempster-Shafer theory, accessed via the evidence pane). One will be able to control the size of the set of inferred culprits by setting thresholds on matches and restricting evidence to just some kinds of evidence (e.g., fingerprint or mugshot). To find details of an alleged culprit, one can follow up with the suspect template mentioned above.

The tab for classifying the case and its constituent situations provides an interface for applying some of the SWRL rules mentioned in Section 6. We could find the classes (if any) that the various situations instantiate. For example, a situation might be an instance of the abstract id-situation where an attempt is made to identify a culprit by fingerprint. We could find the id-case class (if any) that an id-case instantiates. For example, a given constellation of situations forming an id-case may be an instance of a case where a culprit is identified by fingerprint. Finally, we could determine whether a given id-case is coordinated with another id-case as they involve the same described situation and similar set of suspects.

9.3 Evidence Pane

The evidence pane will support the application of Dempster-Shafer theory to the evidence provided for the cases. It is assumed that mass functions have been defined for each kind of evidence in each case. Like the query pane, the contents of the evidence pane will be specific to the selected case. The user will be able to access the templates for suspects and for personnel in this pane as well to see details on people of interest.

There will be a tab for each kind of evidence and a tab for the combined evidence. The contents of the tabs for the various kinds of evidence will have similar structure. Consider, for example, the tab for the fingerprint evidence. The user will be able to request a table for the singleton sets of suspects ordered by belief or ordered by plausibility, given the mass function based on fingerprint matches. This could be restricted to the top few in belief or plausibility. Frequently, the belief or plausibility of non-singleton sets is of interest, particularly when the belief for such a set is high or the plausibility is low. The user might request small sets with high belief or large sets with low plausibility. The mass functions can me modified by features of the resource situations, such as the reliability of the forensic professional who took the fingerprint on file. The fingerprint tab will include ways for the user to have such modifying aspects incorporated into the mass function.

The tab for the combined evidence will allow the user to view tables of singleton sets, now with the belief and plausibility from the combined mass function. Non-singleton sets are of interest here as they are with a single kind of evidence. The user will be able to select the combination rule used, the default being Dempster's rule. There will also be a way to analyze sensitivity. For example, the fingerprint from the scene might be quite indistinct so that it does not discriminate sharply between the suspects while the forensic mugshot may be a clear picture of the culprit's face. In that case, the page should indicate that identification relies much more on the one kind of evidence (mugshots) than on the other (fingerprints) and give some indication of how much more.

10 Implementation of the Web-based System

This section describes the implementation of our web-based system, which is a work in progress. To implement the backend of our system, we used the Apache Struts model-view-controller framework (The Apache Software Foundation 2010), Apache Velocity template engine (The Apache Software Foundation 2014), and the Stardog triplestore [12]. The view (frontend) and the controller interface are provided by Struts and can integrate with other technologies to provide the model. A controller acts as a bridge between an application's model and the web view. The Stardog triple store (Complexible 2014) supports OWL and rule reasoning, which it does in a lazy and late-binding fashion: reasoning is performed at query time according to a reasoning type specified by the user. Apache Velocity is a template engine for Java that provides the user a simple but powerful template language able to reference objects defined in Java. We use Velocity to create SPARQL queries from a template since we assume that the user is not familiar with SPARQL. We connect to the Stardog triplestore using the RESTful web service exposed by Stardog. We decided to use the RESTful web services because the Jena (Apache Software Foundation 2013) interface code was not available. (Jena is our preferred Semantic Web framework.) A RESTful web service uses HTTP verbs for accessing and manipulating data.

To keep the web application running as fast as possible, an XML document was created to store meta-information about and control information for our scenarios. The information in this XML document includes the id, name, and description of the scenario. The user will access the data in this XML document through the interface. The XML document will also contain the possible queries that can be executed on each piece of evidence as well as the types of evidence for each scenario. The code will turn similarity measures into masses used later for belief and plausibility calculations.

Belief and plausibility calculations are written in Python. To run this script from Java, we use the Jython interpreter (Jython 2001), which is a Python interpreter written in Java and is embeddable in Java applications. To store the data produced by the code executed by Jython, a data structure was created to store belief and the plausibility values. The data structure with the masses will be stored into a session variable to be used later. After the user selects the combination rule to use, the data structure will be passed into a combination method with a string identifier that identifies the combination rule. The Jython interpreter will run the appropriate method for that particular combination rule. The output of the running of the interpreter will be stored into a data structure of data objects. The data produced by the Jython interpreter will be cached into memory using Infinispan (Redhat 2009), which is an in-memory key-value data store written in Java and used as a cache or a data grid. It can be embedded into Java applications or used as a remote service over a variety of protocols. Our setup will have Infinispan embedded into our application so that some of the data may be cached .

11 Conclusion

We have presented our computational framework for agent identity, currently focused on criminal investigations, especially the use of biometrics therein. This paper is especially concerned with ongoing development of a web-based system that makes available via a webpage the functionality of this framework especially for pedagogical putposes. The theoretical underpinnings are in Barwise's situation theory. We construct a case as a constellation of situations, including an id-situation, corresponding to Barwise's utterance situation and involving an identity assertion (of the culprit). The described situation is the crime scene. There are also resource situations tied by convention to the id-situation, such as where fingerprints are recorded and filed. Cases are encoded in RDF, and several OWL ontologies have been developed to serve as a knowledge base and define RDF classes and properties. Dempster-Shafer theory is used to handle uncertainty and in combining levels of possibly conflicting or corroborating evidence. We presented an example and discussed its encoding in RDF as per our ontologies. We discussed our ontologies and the SWRL rules that enhance them. Evidence in the legal sense was discussed in the current context, and the importance of biometric artifacts that persist across situations was noted. A functional design of our webpage was presented, and the ongoing implementation of the web-based system was discussed.

The webpage will make available to students of criminal justice our work on a computational framework on identity as it is particularly focused on identifying the culprit in a crime scene investigation. A computerized evidentiary fact pattern would help students develop their critical thinking skills and support interdisciplinary problem solving. Additionally, the availability of the web interface will provide unlimited opportunities for students to practice and engage in fact-pattern scenarios, thus building valuable experience for future careers and research. The Criminal Justice program at North Carolina A&T State University will use the webpage in four of its courses and in its signature co-curricular project. The webpage can be used as an instructional tool and for formative assessment. The web interface could also be used in the Aggie Sleuths Project, an interdisciplinary, interdepartmental research project based on a simulated crime scene (Fakayode et al. 2016).

We have encoded several scenarios besides the example given in this paper, and we are in the process of encoding several more. Note that fingerprint and mugshot biometrics are used online. In working with online authentication (Jenkins et al.), we are addressing behavioral biometrics (e.g., swipe patterns on mobile devices). We intend eventually to address any kind of evidence for identity and to develop ontologies as required. And we shall continue to enhance our use of Dempster-Shafer theory, looking at various combination rules and ways to modify mass functions.

References

Apache Software Foundation. 2013. Apache jena.

Barwise, J., and Perry, J. 1983. *Situations and attitudes*. Center for the Study of Language and Information.

Barwise, J. 1989. *The situation in logic*. Center for the Study of Language (CSLI).

Black, H. C.; Garner, B. A.; and McDaniel, B. R. 1999. *Black's law dictionary*. West Group St.

Citizens Information. 2014. Real evidence.

Complexible. 2014. Stardog.

Creese, S.; Gibson-Robinson, T.; Goldsmith, M.; Hodges, D.; Kim, D.; Love, O.; Nurse, J. R.; Pike, B.; and Scholtz, J. Tools for understanding identity. In *Technologies for Homeland Security (HST), 2013 IEEE International Conference on*, 558–563. IEEE.

Devlin, K. 1995. *Logic and information*. Cambridge University Press.

Fakayode, S. O.; Mayes, J. P.; Kanipes, M. I.; Johnson, D.; and Cuthbertson, E. L. 2016. Promoting student learning in criminal justice, stem, and forensic science: Aggie sleuth initiative (aggiesi)-guided inquiry learning experience. *Journal of Criminal Justice Education* 1–15.

Findlaw. 2016. Summary of the rules of evidence.

Galton, A., and Mizoguchi, R. 2009. The water falls but the waterfall does not fall: New perspectives on objects, processes and events. *Applied Ontology* 4(2):71–107.

Giannelli, P. C. 1993. Chain of custody. Report, School of Law, Case Western Reserve University.

Halpern, J. Y. 2005. *Reasoning about uncertainty*. MIT Press.

Jenkins, J.; Nick, W.; Roy, K.; Esterline, A.; and Bloch, J. Author identification using sequential minimal optimization. In *SoutheastCon, 2016*, 1–2. IEEE.

Jython. 2001. Jython: Python for the java platform.

McDaniel, M.; Sloan, E.; Day, S.; Mayes, J.; Nick, W.; Roy, K.; and Esterline, A. 2017a. Situation-based ontologies for a computational framework for identity focusing on crime scenes. In *Accepted for IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA 2017)*.

McDaniel, M.; Sloan, E.; Nick, W. Mayes, J.; and Esterline, A. 2017b. Ontologies for situation-based crime scene identities. In *Accepted for IEEE SoutheastCon*.

Pan, J. Z. 2009. Resource description framework. In *Handbook on Ontologies*. Springer. 71–90.

Pérez, J.; Arenas, M.; and Gutierrez, C. 2006. Semantics and complexity of sparql. In *International semantic web conference*, 30–43. Springer.

Redhat. 2009. Infinispan.

Sloan, E.; McDaniel, M.; Nick, W.; Mayes, J.; and Esterline, A. 2016. Structure and evidence in identity cases. In *Proc. 2016 IEEE Symposium Series on Computational Intelligence (IEEE SSCI 2016)*.

Sloan, E.; McDaniel, M.; Nick, W.; Mayes, J.; and Esterline, A. 2017. Marshalling situation-based evidence in identity cases. In *Accepted for IEEE SoutheastCon*.

The Apache Software Foundation. 2010. Apache struts.

The Apache Software Foundation. 2014. Apache velocity.