

## Knowledge provenance in enterprise information

M. S. FOX\* and J. HUANG

Enterprise Integration Laboratory, University of Toronto, 40 George Street,  
Toronto, Ontario M5S 3G8 Canada

(Revision received January 2005)

Knowledge Provenance (KP) addresses the problem of how to determine the validity and origin of web information by means of modelling and maintaining information sources, information dependencies, and trust structures. Four levels of KP have been identified: (1) static KP develops the fundamental concepts for KP, and focuses on provenance of static and certain information; (2) dynamic KP considers how the validity of information may change over time; (3) uncertainty-oriented KP considers uncertain truth values and uncertain trust relationships; (4) judgement-based KP focuses on social processes necessary to support KP. This paper presents the fundamental concepts and models of KP by providing motivating scenarios, KP ontologies and examples regarding how to use KP.

**Keywords:** Enterprise integration; Knowledge management;  
Information trustworthiness; Knowledge Provenance; Web of trust

### 1. Introduction

The information revolution continues unabated throughout the Enterprise. The confluence of broadband global communications, inexpensive computers, information standards such as the World Wide Web and visualization technologies such as web browsers has made it possible to acquire, store, refine and disseminate vast quantities of information throughout the enterprise. Consequently, web-based Enterprise Portals have fast become the primary means by which people find and publish information across the enterprise. But there is a problem lurking behind the portal, and it arises from many directions: information may no longer be relevant (e.g. discontinued products, old operating procedures, old financial information), may contain incorrect information (e.g. news stories), and may even be outright lies. Anyone can publish information on the web, the information may be true or false, valid or dated, but no tool exists to discern the differences.

This paper introduces Knowledge Provenance (hereafter, referred as KP) (Fox and Huang 2003) to create an approach to determining the origin and validity of web information by means of modelling and maintaining information sources, information dependencies, and trust structures. The major questions KP attempts to answer include: Can this information be believed to be true? Who created it? Can its creator

\*Corresponding author. Email: msf@eil.utoronto

be trusted? What does it depend on? Can the information it depends on be believed to be true? This proposed approach could be used to help people and web software agents to determine the validity of web information.

Philosophically, we believe the web will always be a morass of uncertain and incomplete information. But we also believe that it is possible to annotate web content to create islands of certainty. Towards this end, KP introduces four levels of provenance that range from strong provenance (corresponding to high certainty) to weak provenance (corresponding to high uncertainty). Level 1 (static KP) (Fox and Huang 2003) develops the fundamental concepts for KP and focuses on provenance of static and certain information; Level 2 (dynamic KP) (Huang and Fox 2004a) considers how the validity of information may change over time; Level 3 (uncertainty-oriented KP) (Huang and Fox 2004b) considers uncertain truth value and uncertain trust relationships; Level 4 (judgement-based KP) focuses on social processes necessary to support KP.

The content of this paper is organized as follows: section 2 provides an overview of related research; section 3 introduces an Ontology for static KP; section 4 illustrates how to use KP including annotating web documents, provenance reasoning and implementations; section 5 extends static KP into dynamic KP; section 6 further extends static KP into uncertainty-oriented KP; finally, section 7 concludes the paper.

## 2. Related research

In the context of enterprise integration, Goranson *et al.* (2002) stated that 'knowledge' in Knowledge Management is 'justified true belief', and two types of trust may be involved: 'deductive' trust based on understanding the cause and effect mechanics; 'inductive' trust based on 'authority', 'votes', and 'experience'. 'Deductive' trust is preferred, for it is 'auditable'. This argument raised an important point: knowledge in EI needs to be justified and to be auditable. Consider the features of the knowledge in EI: uncertain, complex, incomplete, inaccurate, qualitative, informal, and so forth. 'Trust' is a necessary and efficient approach to knowledge justification, and a combination of 'inductive' and 'deductive' trust is needed.

Information source evaluation criteria, such as, Authority, Accuracy, Objectivity, Currency and Coverage, have been developed in library and information science, and have been extended to online information (Alexander and Tate 1999). Oliver (1997) collected hundreds of evaluation criteria from different sources, and consolidate into 125 indicators in 11 groups of criteria. However, these criteria were proposed for people to judge web information, that is, people have to be involved in judgement. KP attempts to create a computational model.

Interest in addressing the issue of web-information trustworthiness has appeared under the umbrella of the 'Web of Trust' that is identified as the top layer of the Semantic Web (Berners-Lee 2003, slides 26 and 27).

Digital signature and digital certification (Simon *et al.* 2001, Tan 2002, Bartel *et al.* 2002) play important roles in 'Web of Trust'. However, they only provide an approach to validate author identification and information integrity; they do not determine information trustworthiness. In the context of KP, they can only be used to determine who is the information creator and whether the information is the same

as originally defined, but they do not indicate whether the information is trustworthy. Trustworthiness is supposed to be evaluated by each web application. For the purposes of secure web access control, Blaze *et al.* (1996) first introduced 'decentralized trust management' to separate trust management from applications and created the fundamental concepts of policy, credential, and trust relationship. Chu (1997) introduced trust protocol in the REFEREE system; Khare and Rifkin (1997) proposed basic principles of trust management. However, Trust Management focuses on 'is this individual trusted to do a specific operation in my system', whereas KP attempts to answer 'is the information given by this individual trusted to be true?'

The concept of 'Web of Trust' perhaps was first developed in PGP (Zimmermann 1995), a public key cryptosystem used for encryption and digital signature, as a model of trust for a public key recipient to validate the authentication of the key, where 'trust' has a specific meaning of trusting people to validate other people's certificates. The 'Friend Of A Friend' project (FOAF, Dumbill 2002) creates social networks (Watts 1999) on the web by facilitating people to describe acquaintance relationships in machine-readable webpages. Even though acquaintance relationships do not equal trust relationships, FOAF is a good start point.

Many projects focusing on 'web of trust' based on social networks have emerged. For example, Yu and Singh (2000) proposed a model of reputation (trust) propagation and building among agents in electronic communities; Golbeck *et al.* (2002) proposed trust networks that extend the FOAF model by introducing levels of trust in acquaintance relationships and used for filtering emails; Richardson *et al.* (2003) proposed a model of trust management for the semantic web and used for bibliography recommendation. These projects created models for indirect trust calculating and could be used for justifying knowledge in the manner of 'inductive' trust.

Coming from an automated reasoning perspective, McGuinness and Pinheiro da Silva (2003) developed 'Inference Web (IW)', which enables information creators to register proofs with provenance information in IW, and then IW is able to explain the provenance of a piece of requested knowledge. IW provides provenance information (registered by creators) for users to support them deciding by themselves to trust or not trust the requested knowledge.

Finally, technologies developed in AI, such as the Truth Maintenance System (de Kleer *et al.* 1989) and Temporal Logic (Allen and Ferguson 1994), provide approaches for knowledge representation and reasoning in KP. Technologies developed in Semantic Web (Berners-Lee *et al.* 2001) provide approaches to the web implementation of KP.

### 3. Static KP

Static KP is concerned with the provenance of knowledge that is both certain and does not change over time. It is the simplest yet strongest form of provenance. Basically, any statement has a truth value of: True, False or Unknown. The default truth value is 'Unknown'. Its truth value does not change over time.

In order to give a formal and explicit specification for Static KP and to make it available on the web, a static KP ontology is defined in this section. Following the ontology development methodology of Gruninger and Fox (1995), we specify Static

KP ontology in four steps: (1) provide a motivating scenario; (2) define informal competency questions for which the ontology must be able to derive answers; (3) define the terminology (i.e. predicates); (4) define the axioms (i.e. semantics). We already discussed motivating scenarios in the earlier section. This section presents informal competency questions, terminology, and axioms.

### 3.1 *Motivating scenarios*

Case 1 (Asserted information): Consider a statement in a document found on the Intranet in an enterprise. This states that 'a delay of more than one minute in answering a phone call may result in the customer being dissatisfied'. From a provenance perspective, there are three questions that have to be answered:

- Is the statement true?
- Who created this statement?
- Should we believe the person or organization that created it?

A further examination of the text of the web document provides the answers: it can be believed to be true, created by a retired customer service manager, who most people in the company believe is an authority on the subject. Questions are:

- What is the basis for us to believe this statement as true?
- How can the provenance process be formalized?

In the remainder of the paper, we will refer to statements like the above as 'propositions'. A proposition is the smallest piece of information to which provenance-related attributes may be ascribed. A proposition is either true or false.

Case 2 (Dependent information): Consider the following proposition found in another web document: 'This new approach to reduce response-delay to less than one minute may increase customer loyalty, because a delay of more than one minute in answering a phone call may cause the customer to be unsatisfied'. This is actually two propositions composed of a premise, 'A delay of more than one minute in answering a phone call may cause the customer to be unsatisfied' and a conclusion, 'This new approach to reduce response-delay to less than one minute may increase customer loyalty'. Just as in the previous case, the same questions need to be answered for each proposition. What makes this case more interesting is that answering these questions is dependent upon propositions found in other web pages. There are two types of dependency occurring. First, the truth of the premise is dependent on the truth of the proposition found in another web document. Second, the truth of the conclusion depends on the truth of the premise and upon some hidden reasoning that led to the deduction. These types of propositions are called 'dependent propositions' in KP.

It is common to find information in one document that is reproduced in another. The reproduction of a proposition in a second document leads to an equivalence relationship between the two propositions, i.e. the truth values of the two propositions are equivalent. However, the relationship is also asymmetric; one proposition is a copy of the other. The copy of one proposition is classified as 'equivalent proposition'. Furthermore, a proposition can be derived using logical deduction. Hence, the truth value of the derived proposition depends on the truth values of its antecedent propositions. This type of derived proposition is classified as 'derived proposition'.

Returning to the example, determining the provenance of the premise requires that we link, in some way, the premise to the proposition in the other web document from which it is copied. The same is true of the conclusion. Minimally, we should link it to its premise, maximally we should link it to the axioms that justify its derivation. These links will also require some type of certification so that we know who created it and whether it is to be trusted.

From these two cases, a number of concepts required for reasoning about provenance emerge:

- Text is divided into propositions. Once so designated, they are assumed to be indivisible.
- A proposition must have a digital signature so that we can guarantee the identity of the creator.
- An assertion is believed to be true, if the information user trusts the person or organization that created the assertion in the corresponding topic.
- As propositions are reused across the web, a link between where it is used and where it came from must be maintained. These links, or dependencies, must be included in the digital signatures with propositions.
- Dependencies can be simple copies or can be the result of a reasoning process. If the latter, then axioms used in the reasoning should also be identified and signed by an acceptable organization.

### 3.2 Informal competency questions

The competency of an ontology is defined by a set of questions. In other words, the ontology contains the terms and axioms necessary to answer the competency questions. The competency of Static KP ontology is illustrated by the following questions:

- Is this proposition true, false, or unknown?
- Who created this proposition?
- What is the digital signature verification status?
- Which knowledge fields does this proposition belong to?
- In these fields, can the information creator be trusted?
- Does the truth of this proposition depend on any other propositions? If so, what?

### 3.3 Terminology

There are five main classes in static KP ontology: Propositions, Documents, Information Sources, Trust Relationships and Signature Status. From the motivating scenario in section 3.1 and the nature of propositions, a taxonomy of propositions in the Static KP Ontology is constructed as shown in figure 1.

**3.3.1 Propositions.** KP-Prop is the most general concept used to represent propositions in a document. Table 1 defines the predicates for depicting a KP proposition and its attributes.

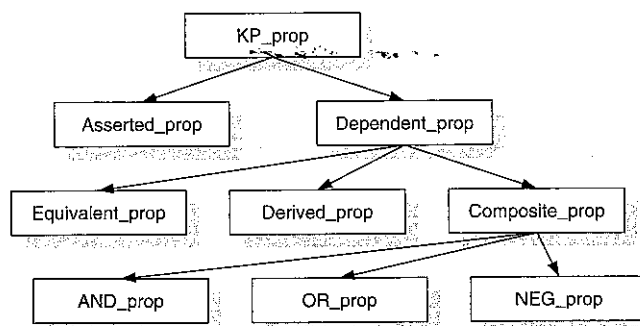


Figure 1. Proposition taxonomy in KP.

Table 1. Proposition-related predicate definitions in static KP ontology.

Predicate	Description
<i>type</i> ( <i>x</i> , 'KP_prop')	<i>x</i> is defined to be a proposition, signified by being of type KP_prop
<i>proposition_content</i> ( <i>x</i> , <i>s</i> )	<i>s</i> is the content of the proposition <i>x</i> ; In html files, the content of a proposition usually is a string; in xml files, the content of a proposition can be an xml element
<i>assigned_truth_value</i> ( <i>x</i> , <i>v</i> )	Proposition <i>x</i> has a truth value <i>v</i> assigned by proposition creator.
<i>trusted_truth_value</i> ( <i>a</i> , <i>x</i> , <i>v</i> )	Agent <i>a</i> trusts that proposition <i>x</i> has a truth value <i>v</i> . <i>v</i> may be one of 'True', 'False', or 'Unknown'
<i>type</i> ( <i>x</i> , 'asserted_prop')	<i>x</i> is an assertion and not dependent upon any other proposition
<i>type</i> ( <i>x</i> , 'dependent_prop')	<i>x</i> is a proposition whose truth value is dependent upon another proposition; dependent-prop class is further divided into three subclasses: equivalent-prop, derived-prop, and composite-prop
<i>type</i> ( <i>x</i> , 'equivalent_prop')	An equivalent-prop is a copy of and its truth value is the same as the proposition it depends on.
<i>type</i> ( <i>x</i> , 'composite_prop')	Composite-prop's is defined to be the logical combination of its constituent propositions; a composite-prop is divided into three subclasses: neg-prop, and-prop, and or-prop
<i>type</i> ( <i>x</i> , 'derived_prop')	A derived-prop indicates that the proposition is a derived conclusion based on some premises; for example, derived-prop B has dependency-link pointing to composite-prop A, which means that A is a premise of B
<i>is_dependent_on</i> ( <i>x</i> , <i>y</i> )	Proposition <i>x</i> is dependent on proposition <i>y</i> . <i>x</i> is called the dependent proposition, and <i>y</i> is called the support proposition
<i>has_same_content</i> ( <i>x</i> , <i>y</i> )	Proposition <i>x</i> has the same proposition content as <i>y</i>

**3.3.2 Documents.** To facilitate the determination of the provenance of a proposition, properties of the document in which it appears may need to be considered. For example, knowing who created the document may be important in determining the validity of a proposition within. A document can be any type of file. For the purposes of this paper, we restrict our attention to standard web files such as html files, xml files and xhtml files. Document-related KP predicates are defined in table 2.

Predi

has\_in

has\_si

has\_si

Predi

trusted

trusted

in\_field  
subfield3.3.3 I  
creator  
status  
software  
predica3.3.4 T  
context  
proposi  
where tl  
c in a t  
proposi  
predicat

3.4 Axi

In the fc  
Basically  
propositi

Table 2. Document-related predicate definitions in static KP ontology.

Predicate	Definition
$type(x, 'document')$	$x$ is defined to be a KP document
$in\_document(y, d)$	Proposition $y$ is contained in document $d$

Table 3. Information source-related predicate definitions in static KP ontology.

Predicate	Description
$has\_infoCreator(x, c)$	KP-prop or Document $x$ has infoCreator $c$ ; here, infoCreator may be either creator or publisher
$has\_signature(x, s)$	The proposition or document $x$ has a signature $s$
$has\_sig\_status(x, v)$	The digital signature verification status of $x$ is $v$ , where $v$ may be one of three status: 'Verified'—the signature is verified successfully; 'Failed'—the signature verification is failed; and 'NoSignature'—does not have digital signature

Table 4. Trust-related predicate definitions in static KP ontology.

Predicate	Description
$trusted\_in(a, c, f)$	Provenance requester $a$ (directly or indirectly) trusts information creator $c$ in knowledge field $f$
$trusted(x, a)$	Proposition $x$ is trusted by agent $a$ ; this means its information creator is trusted by $a$ in one of the fields which proposition $x$ belongs to
$in\_field(x, f)$	Proposition $x$ belongs to knowledge field $f$
$subfieldOf(x, y)$	Knowledge field $x$ is a sub-field of knowledge field $y$

**3.3.3 Information source and signature.** For any document and proposition, its creator can be defined. Along with this, a digital signature and the verification status of the signature can be defined. Assume that a digital signature validation software provides the result of signature verification. Table 3 defines the related predicates.

**3.3.4 Trust relationships.** Earlier we stated that KP is context-sensitive, where the context is trust relations that define whether the provenance requester trusts whose propositions in what topics. A trust relationship in KP is defined as a triple  $(u, c, f)$  where the provenance requester (information receiver)  $a$  'trusts' information creator  $c$  in a topic or a specific knowledge field  $f$ ; here, 'trust' means that  $a$  believes any proposition created by  $c$  in field  $f$  to be true. Table 4 defines the trust-related predicates.

### 3.4 Axioms

In the following, a set of axioms is defined to specify truth conditions of KP-props. Basically, the truth value of an asserted proposition depends on whether the proposition is 'trusted'; the truth value of an equivalent proposition depends on

the truth value of its support KP-prop that the equivalent proposition points to by its dependency link; the truth value of a derived proposition depends on whether the proposition is 'trusted' and whether its support KP-prop is true. In addition, a KP-prop is 'trusted' if the creator or publisher of the proposition is trusted in one of the fields of the proposition, and the digital signature verification status is 'Verified'. Finally, the 'close world assumption' is applied to handle 'not' in this paper.

**3.4.1 Asserted propositions.** An asserted-prop is trusted to have its truth value as assigned, if the asserted-prop is trusted by the provenance requester.

Axiom SKP-1:

$$\text{for-all } (a, x, v) ((\text{type}(x, \text{'asserted\_prop'}) \wedge \text{trusted}(x, a) \wedge \text{assigned\_truth\_value}(x, v)) \rightarrow \text{trusted\_truth\_value}(a, x, v)).$$

A KP-prop is 'trusted', if the creator or publisher of the proposition is 'trusted' in one of the fields of the proposition, and the digital signature verification status is 'Verified'.

Axiom SKP-2:

$$\text{for-all } (a, x, f, c, w) ((\text{type}(x, \text{'KP-prop'}) \wedge \text{has\_sig\_status}(x, \text{'Verified'}) \wedge \text{has\_infoCreator}(x, c) \wedge \text{in\_field}(x, f) \wedge \text{trusted\_in}(a, c, w) \wedge \text{subfield\_of}(f, w)) \rightarrow \text{trusted}(x, a)).$$

For a KP-prop that has no creator specified, the creator of the document is the default creator of the KP-prop.

Axiom SKP-3:

$$\text{for-all } (x, d, c) ((\text{type}(x, \text{'KP-prop'}) \wedge (\text{not}(\text{exist}(c2) \text{has\_creator}(x, c2))) \wedge \text{in\_document}(x, d) \wedge \text{has\_creator}(d, c)) \rightarrow \text{has\_creator}(x, c)).$$

If a proposition does not have a creator, then the digital signature verification status of the KP-prop is determined by the digital signature verification/status of the document.

Axiom SKP-4:

$$\text{for-all } (x, d, c, v) ((\text{type}(x, \text{'KP-prop'}) \wedge (\text{not}(\text{exist}(c2) \text{has\_creator}(x, c2))) \wedge \text{in\_document}(x, d) \wedge \text{has\_creator}(d, c) \wedge \text{has\_sig\_status}(d, v)) \rightarrow \text{has\_sig\_status}(x, v)).$$

**3.4.2 Equivalent propositions.** The trusted truth value of an equivalent-prop is the same as the trusted truth value of its support proposition, if this equivalent-prop has exactly the same proposition-content as its support proposition has.

Axiom SKP-5:

$$\text{for-all } (a, x, y, v) ((\text{type}(x, \text{'equivalent\_prop'}) \wedge \text{is\_dependent\_on}(x, y) \wedge \text{has\_same\_content}(x, y) \wedge \text{trusted\_truth\_value}(a, y, v)) \rightarrow \text{trusted\_truth\_value}(a, x, v)).$$



**3.4.3 Composite propositions.** The trusted truth value of a neg-prop is the negation of the trusted truth value of the KP-prop it is dependent on.

Axiom SKP-6:

$$\begin{aligned} & \text{for-all } (a, x, y) ((\text{type}(x, \text{'neg\_prop'})} \\ & \wedge \text{is\_dependent\_on}(x, y) \wedge \text{trusted\_truth\_value}(a, y, \text{'True'})) \\ & \rightarrow \text{trusted\_truth\_value}(a, x, \text{'False'})). \end{aligned}$$

Axiom SKP-7:

$$\begin{aligned} & \text{for-all } (a, x, y) ((\text{type}(x, \text{'neg\_prop'})} \\ & \wedge \text{is\_dependent\_on}(x, y) \wedge \text{trusted\_truth\_value}(a, y, \text{'False'})) \\ & \rightarrow \text{trusted\_truth\_value}(a, x, \text{'True'})). \end{aligned}$$

The trusted truth value of an and-prop is 'True' if all its support KP-props are 'True'; and the trusted truth value of an and-prop is 'False' if at least one of its support KP-props is 'False'.

Axiom SKP-8:

$$\begin{aligned} & \text{for-all } (a, x) ((\text{type}(x, \text{'and\_prop'})} \\ & \wedge \text{for-all } (y) (\text{is\_dependent\_on}(x, y) \rightarrow \text{trusted\_truth\_value}(a, y, \text{'True'}))) \\ & \rightarrow \text{trusted\_truth\_value}(a, x, \text{'True'})). \end{aligned}$$

Axiom SKP-9:

$$\begin{aligned} & \text{for-all } (a, x) ((\text{type}(x, \text{'and\_prop'})} \\ & \wedge (\text{exist } (y) (\text{is\_dependent\_on}(x, y) \wedge \text{trusted\_truth\_value}(a, y, \text{'False'})))) \\ & \rightarrow \text{trusted\_truth\_value}(a, x, \text{'False'})). \end{aligned}$$

The trusted truth value of an or-prop is 'True' if at least one of its support KP-props is 'True'; and the trusted truth value of an or-prop is 'False' if all its support KP-props are 'False'.

Axiom SKP-10:

$$\begin{aligned} & \text{for-all } (a, x) ((\text{type}(x, \text{'or\_prop'})} \\ & \wedge (\text{exist } (y) (\text{is\_dependent\_on}(x, y) \wedge \text{trusted\_truth\_value}(a, y, \text{'True'})))) \\ & \rightarrow \text{trusted\_truth\_value}(a, x, \text{'True'})). \end{aligned}$$

Axiom SKP-11:

$$\begin{aligned} & \text{for-all } (a, x) ((\text{type}(x, \text{'or\_prop'})} \\ & \wedge (\text{for-all } (y) (\text{is\_dependent\_on}(x, y) \wedge \text{trusted\_truth\_value}(a, y, \text{'False'})))) \\ & \rightarrow \text{trusted\_truth\_value}(a, x, \text{'False'})). \end{aligned}$$

**3.4.4 Derived propositions.** The trusted truth value of a derived proposition is 'True' or 'False' as specified, if it is 'trusted' and its support KP-prop (condition) is 'True'. Note that the axioms used to derive the truth value do not have to be included as part of the dependency.

Axiom SKP-12:

$$\begin{aligned} & \text{for-all } (a, x, y, v) ((\text{type}(x, \text{'derived\_prop'})} \wedge \text{trusted}(x, a) \wedge \text{assigned\_truth\_value}(x, v) \\ & \wedge \text{is\_dependent\_on}(x, y) \wedge \text{trusted\_truth\_value}(a, y, \text{'True'})) \\ & \rightarrow \text{trusted\_truth\_value}(a, x, v)). \end{aligned}$$

**3.4.5 Default assigned\_truth value.** The default truth value of an asserted or derived proposition assigned by the proposition creator is 'True'.

Axiom SKP-13:

$$\begin{aligned} & \text{for-all } (a, x, y, v) ((\text{type}(x, \text{'asserted\_prop'}) \vee \text{type}(x, \text{'derived\_prop'}) \\ & \quad \wedge \text{triple}(x, \text{assigned\_truth\_value}, v)) \\ & \rightarrow \text{assigned\_truth\_value}(a, x, v)). \\ & \text{for-all } (a, x, y, v) ((\text{type}(x, \text{'asserted\_prop'}) \vee \text{type}(x, \text{'derived\_prop'}) \\ & \quad \wedge \text{not } (\text{triple}(x, \text{assigned\_truth\_value}, v))) \\ & \rightarrow \text{assigned\_truth\_value}(a, x, \text{'True'})). \end{aligned}$$

**3.4.6 Default trusted\_truth value.** The default trusted truth value of a proposition is 'Unknown'.

Axiom SKP-14:

$$\begin{aligned} & \text{for-all } (a, x, v) ((\text{type}(x, \text{'KP\_prop'})} \\ & \quad \wedge \text{not } (\text{trusted\_truth\_value}(a, x, \text{'True'}) \wedge \text{not } (\text{trusted\_truth\_value}(a, x, \text{'False'}) \\ & \rightarrow \text{trusted\_truth\_value}(a, x, \text{'Unknown'})). \end{aligned}$$

#### 4. Implementing KP

In order to apply KP in practice, information creators need to annotate web documents with KP metadata, users (provenance requesters) need to define their trust relationships, and a KP reasoner is needed to conduct provenance reasoning on annotated web documents.

The motivating example introduced in section 3 is used to illustrate how to determine the provenance of propositions as shown in figure 2. In order to facilitate the annotation of web documents with KP metadata and define trust relationships,

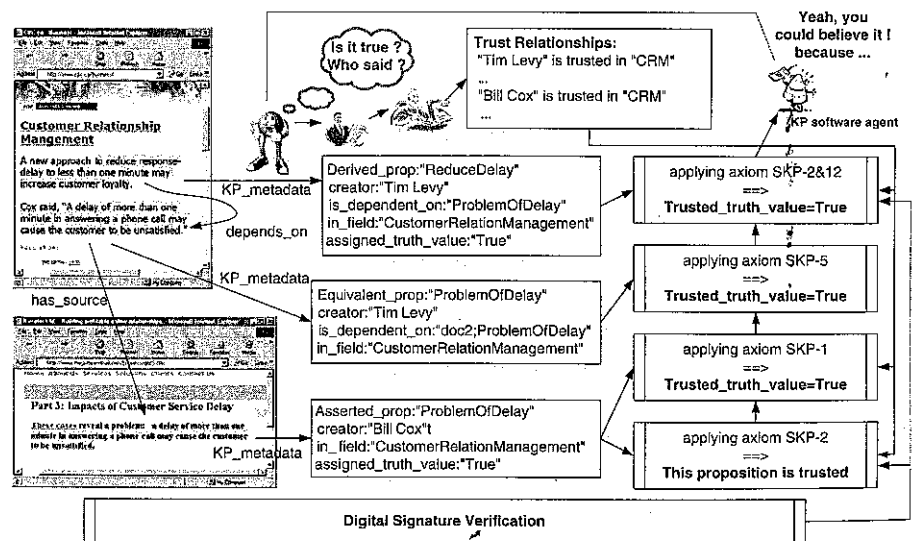


Figure 2. Example to illustrate how to determine the provenance of web information by KP.

n asserted or

we have defined a KP markup language in RDFS (Resource Description Framework Schema; see <http://w3.org/rdf/>). The following is an example of how to annotate a html file with KP metadata.

**Document1:** <http://www.crm-examples.com/ex1.html>

```
< HTML xmlns="http://www.w3.org/1999/xhtml"
dsig="http://www.w3.org/2000/09/xmldsig#"
kp="http://www.eil.utoronto.ca/kp#" >
< HEAD >
< kp:Document rdf:about="http://www.example.com/ex1.html" creator="Tim
Levy"/ >
</HEAD >
< BODY >
< kp:Derived_prop rdf:id="ReduceDelay" is_dependent_on="#ProblemOfDelay"
in_field="Customer Relationship Management" >
The new approach to reduce response-delay to less than one minute may increase
customer loyalty.
</kp:Derived_prop >
< kp:Equivalent_prop rdf:id="ProblemOfDelay"
is_dependent_on="http://www.crm-exam.net/ex2.html#ProblemOfDelay" >
A delay of more than one minute in answering a phone call may cause the customer
to be unsatisfied.
</kp:Equivalent_prop >
< Signature ID="ex1" >
< SignedInfo >
< CanonicalizationMethod Algorithm="http://www.w3.org/TR/2001/REC-xml-
cl4n-20010315"/ >
< SignatureMethod Algorithm="http://www.w3.org/2000/09/xmldsig#dsa-sha1"/ >
< Reference URI="#ReduceDelay" >
< DigestMethod Algorithm="http://www.w3.org/2000/09/xmldsig#sha1"/ >
< DigestValue > j6hm43k9j3u5903h4775si83 </DigestValue >
</Reference >
< Reference URI="#ProblemOfDelay" >
< DigestMethod Algorithm="http://www.w3.org/2000/09/xmldsig#sha1"/ >
< DigestValue > g79lk20rjf023rr032kr93kjr </DigestValue >
</Reference >
</SignedInfo >
< SignatureValue > M459ng9784t... </SignatureValue >
< KeyInfo >
< X509Data >
< X509SubjectName > ... </X509SubjectName >
< X509Certificate > MIID5jCCA0+gA...IVN </X509Certificate >
</X509Data >
< KeyInfo >
</Signature >
</BODY >
</HTML >
```

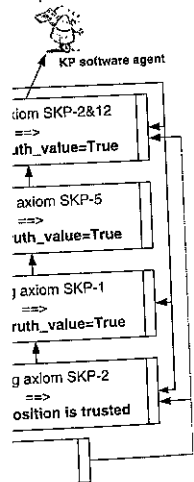
f a proposition

'False')))

otate web docu-  
define their trust  
ce reasoning on

llustrate how to  
rder to facilitate  
ust relationships,

Yeah, you  
could believe it!  
because ...



b information by KP.

In this sample document, there are two KP-props: (1) a 'derived-prop' with id 'ReduceDelay', entailed by KP-prop 'ProblemOfDelay' and (2) an 'equivalent-prop' with id 'ProblemOfDelay'. Let us assume that the digital signature (in XML-Signature syntax) of these two propositions is verified successfully, and Tim Levy is trusted in the field of 'Customer Relationship Management' by the person requesting KP. Therefore, according to Axiom SKP-2, the derived-prop 'ReduceDelay' will be trusted. But to determine whether its truth value is to be trusted, we have to determine whether the support proposition it depends on has a trusted\_truth\_value of true. The proposition 'ProblemOfDelay' is an equivalent proposition, its truth depends on its source, another proposition in another web document: 'http://www.crm-examp.net/ex2.html#ProblemOfDelay'.

*Document2: http://www.crm-examples2.net/ex2.html*

```
<HTML xmlns="http://www.w3.org/1999/xhtml"
dsig="http://www.w3.org/2000/09/xmldsig#"
kp="http://www.eil.utoronto.ca/kp#" >
<HEAD>
<kp:Document rdf:about="http://www.crm-exam.net/ex2.html#ProblemOfDelay">
<kp:creator> "Bill Cox" </kp:creator>
</kp:Document>
</HEAD>
<BODY>
<kp:asserted_prop rdf:id="ProblemOfDelay" in_field="Customer Service">
A delay of more than one minute in answering a phone call may cause the customer
to be unsatisfied.
</kp:asserted_prop>
<Signature ID="Bill-ProblemOfDelay"> ... </Signature>
</BODY>
</HTML>
```

In document 2, 'ProblemOfDelay' is an 'asserted-prop'. Its trusted truth value depends on whether it is trusted. The creator of the document is 'Bill Cox'. Assume the digital signature is verified successfully, and the requestor trusts Bill in the field of 'Customer Service'. Then, according to axioms SKP.2 and 1, 'ProblemOfDelay' has a trusted\_truth\_value of 'True'. Consequently, the equivalent proposition 'ProblemOfDelay' in document 1 also has a trusted truth value of 'True' by using axiom SKP-5, and finally, the derived proposition 'ReduceDelay' has a trusted truth value of true by using axioms SKP-2 and 12. We have implemented the KP reasoner with Prolog. The system can infer the truth of any KP-prop.

## 5. Dynamic KP

This section extends static KP into Dynamic KP to address the problem of how to determine the truth values of web propositions that change over time.

Table 5. Predicate definitions for dynamic KP ontology.

Predicate	Description
$Effective\_from(x, t_1)$	$x$ is effective from time point $t_1$
$Effective\_to(x, t_2)$	$x$ is effective till time point $t_2$
$effective\_at(x, t)$	KP_prop or trust relationship $x$ is effective at time point $t$

### 5.1 Motivating scenario

In the 'real world', the truth value of a proposition may change over time. Consider the supply chain, the prices of products change over time, inventory changes over time, warehouse space changes over time, etc. For example, a computer retailer receives a proposition from its CPU supplier stating 'Intel P4 3.06 GHz Processor is in stock' (valid from 20 to 26 January 2003), which is true in the specified period, but may be false before or after the period. Furthermore, a trust relationship may change over time also, that is, an information creator may be trusted only within a specified period. In the example above, assume there is a contract between the retailer and its supplier effective from 1 January 2003 to 31 December 2003. During this period, the reseller trusts product information provided by the supplier to be true. However, if the contract has expired, the retailer will no longer trust the supplier.

These examples reveal that the truth value of a proposition may be effective only in a specified period, and a proposition creator may be trusted in a topic only within a specified period. So, trust relationships further constrain the periods of truth value validity.

### 5.2 Terminology

In the motivating scenario, we found that the truth value of a proposition may be effective only within a specified time period. We call this period the proposition's 'effective period'. When the truth value of a proposition is effective at a given time point, the proposition is called 'effective' at the time point. We also found that a proposition creator may be trusted only within a specified time period. That is, a trust relationship  $(a, c, f)$ , which means provenance agent  $a$  trusting proposition creator  $c$  in field  $f$ , may have an effective period also. A trust relationship is called 'effective' at a given time point, if the time point is within the effective period of the trust relationship.

In order to describe effective period, we define the predicates to be used in table 5. Furthermore, several predicates defined in Static KP need to be extended with time as shown in table 6.

### 5.3 Axioms

The following axioms provide the semantics for Dynamic KP.

At a given time point, if a KP\_prop is not effective at the time point, it has trusted truth value of 'Unknown'.

Axiom DKP-1:

$$\text{for-all } (a, x, t, v) ((\text{type}(x, \text{'KP\_prop'})} \wedge \text{not}(\text{effective\_at}(x, t))) \\ \rightarrow \text{trusted\_truth\_value\_at}(a, x, \text{'Unknown'}, t))$$

Table 6. Extended predicates defined in static KP.

Predicate	Description
$trusted\_in\_during(a, c, f, t_1, t_2)$	Agent $a$ trusts information creator $c$ in knowledge field $f$ from time point $t_1$ to time point $t_2$ ; here, $[t_1, t_2]$ is called the trust relationship effective period in this paper
$trusted\_in(a, c, f, t)$	Agent $a$ trusts information creator $c$ in knowledge field $f$ at time point $t$ , which is also called the trust relationship; $(a, c, f)$ is effective at time point $t$
$trusted(x, a, t)$	Proposition $x$ is trusted by agent $a$ at time point $t$
$trusted\_truth\_value(a, x, v, t)$	Agent $a$ trusts that proposition $x$ has a truth value $v$ at time point $t$

The default effective period is 'indefinitely', that is, if the *effective\_from*/*effective\_to* is not specified, the *effective\_from*/*effective\_to* will be negative infinite (denoted as  $-M$ )/positive infinite (denoted as  $+M$ ).

Axiom DKP-2:

$$\begin{aligned} &for\_all(x)((type(x, 'asserted\_prop') \vee type(x, 'derived\_prop')) \wedge not(exist(t1) \\ &effective\_from(x, t1))) \\ &\rightarrow effective\_from(x, -M)) \end{aligned}$$

Axiom DKP-3:

$$\begin{aligned} &for\_all(x)((type(x, 'asserted\_prop') \vee type(x, 'derived\_prop')) \wedge not \\ &(exist(t1)effective\_to(x, t1))) \\ &\rightarrow effective\_to(x, +M)) \end{aligned}$$

At a given time point, an *asserted\_prop* is effective if the time point is within its effective period.

Axiom DKP-4:

$$\begin{aligned} &for\_all(t, x, t1, t2) \\ &((type(x, 'asserted\_prop') \wedge effective\_from(x, t1) \wedge effective\_to(x, t2) \wedge t1 \leq t \wedge t \leq t2) \\ &\rightarrow effective\_at(x, t)) \end{aligned}$$

At a given time point, a *derived\_prop* is effective if: (1) the time point is within the effective period of the *derived\_prop*; and (2) its support *KP\_prop* is effective at the time point.

Axiom DKP-5:

$$\begin{aligned} &for\_all(t, x, y, t1, t2)((type(x, 'derived\_prop') \wedge is\_dependent\_on(x, y) \wedge effective\_at(y, t) \\ &\wedge effective\_from(x, t1) \wedge effective\_to(x, t2) \wedge t1 \leq t \wedge t \leq t2) \\ &\rightarrow effective\_at(x, t)) \end{aligned}$$

At a given time point, an *equivalent\_prop*/*neg\_prop* is effective if its dependency node is effective at the time point.

Axiom DKP-6:

$$\begin{aligned} &for\_all(x, y, v, t)((type(x, 'equivalent\_prop') \vee type(x, 'neg\_prop')) \\ &\wedge is\_dependent\_on(x, y) \wedge effective\_at(y, t)) \\ &\rightarrow effective\_at(x, t)) \end{aligned}$$

At a given time point, an *and\_prop* is effective, if all its support *KP\_props* are effective at the time point.

Axiom DKP-7:

$$\text{for-all}(x, t) ((\text{type}(x, \text{'and\_prop'}) \wedge \text{for-all}(y) (\text{is\_dependent\_on}(x, y) \rightarrow \text{effective\_at}(y, t))) \rightarrow \text{effective\_at}(x, t))$$

At a given time point, an *or\_prop* is effective, if at least one of its support *KP\_props* is effective at the time point.

Axiom DKP-8:

$$\text{for-all}(x, t) ((\text{type}(x, \text{'or\_prop'}) \wedge \text{exist}(y) (\text{is\_dependent\_on}(x, y) \wedge \text{effective\_at}(y, t))) \rightarrow \text{effective\_at}(x, t))$$

Further, 'trusted' is extended as follows. A *KP-prop* is 'trusted' at a given time point, if the creator or publisher of the proposition is 'trusted' in one of the fields of the proposition at the time point, and the digital signature verification status is 'Verified'.

Axiom DKP-9:

$$\text{for-all}(a, x, f, z, c, w, t) ((\text{type}(x, \text{'KP-prop'}) \wedge \text{has\_infoCreator}(x, c) \wedge \text{in\_field}(x, f) \wedge \text{trusted\_in}(a, c, w, t) \wedge \text{subfield\_of}(f, w) \wedge \text{has\_sig\_status}(x, \text{'Verified'})) \rightarrow \text{trusted}(x, a, t))$$

An information creator is trusted in a specific knowledge field at a given time point, if the time point is within the effective period of the trust relationship.

Axiom DKP-10:

$$\text{for-all}(a, c, f, t, t1, t2) ((\text{trusted\_in\_during}(a, c, f, t1, t2) \wedge t1 \leq t \wedge t \leq t2) \rightarrow \text{trusted\_in}(a, c, f, t))$$

Finally, the axioms regarding 'trusted truth value' are extended with time-related predicates '*effective\_at(x, t)*' and '*trusted(x, a, t)*'. The following is the dynamic *KP* axiom for asserted propositions. At a given time point, an asserted-prop has its truth value as specified, if the asserted-prop is trusted by the provenance agent at the time point, and the proposition is effective at the time point.

Axiom DKP-11:

$$\text{for-all}(a, x, v, t) ((\text{type}(x, \text{'asserted\_prop'}) \wedge \text{trusted}(x, a, t) \wedge \text{effective\_at}(x, t) \wedge \text{assigned\_truth\_value}(x, v)) \rightarrow \text{trusted\_truth\_value}(a, x, v, t))$$

The axioms for other types of propositions are extended in similar way, and details can be found in (Huang and Fox 2004a).

## 6. Uncertainty-oriented KP

It is common to find that a person may trust, rather than completely trust or completely distrust, an information creator to a certain degree; furthermore, the proposition created by the information creator may be believed to be true in a certain degree rather than absolutely 'True' or 'False'.

This section extends our KP model to consider uncertain truth values and uncertain trust relationships. The degree of confidence (subjective probability) for a proposition to be true is introduced to extend the binary truth value to a numerical truth value; the degree of belief (subjective probability) is introduced as a trust degree to represent uncertain trust relationships; and a model to calculate the trusted truth value of a proposition is derived with probability theory and information theory.

First, several notations and definitions used in this section are introduced:

- $Pr(Y)$  denotes the probability of event  $Y$ ;
- ' $TTV_x$ ' denotes  $trusted\_truth\_value(a, x, 'True')$ , that is, the trusted truth value of proposition  $x$  (trusted by KP agent  $a$ ) is 'True';
- ' $ATV_x$ ' denotes  $assigned\_truth\_value(x, 'True')$ , i.e. the truth value of proposition  $x$  assigned by proposition creator is 'True';
- ' $Trusted_x$ ' denotes  $trusted(x, a)$ , that is, KP agent  $a$  trusts proposition  $x$ .

When only one proposition is involved, the subscript representing the proposition may be omitted, such as, ' $TTV_x$ ' is written as ' $TTV$ '.

Consider that a proposition has only two possible truth values: 'True' or 'False'; therefore, ' $\neg ATV_x$ ' represents  $assigned\_truth\_value(x, 'False')$ ; and similarly ' $\neg Trusted_x$ ' denotes that agent  $a$  distrusts proposition  $x$ . Note that as a simple method to handle uncertainty, 'Unknown' was used to represent a status in which the truth value cannot be determined in level 1 and 2. In this section, we will use certainty degree to represent uncertain truth value. So, 'Unknown' will no longer be used.

**Definition 1:** The assigned certainty degree (denoted as  $acd$ ) of a proposition given by the proposition creator is defined as the degree of confidence (subjective probability) of the proposition creator to assign the truth value of 'True' to the proposition. It is used to represent the certainty of the assigned truth value.

$$acd = Pr(ATV). \quad (1)$$

**Definition 2:** The certainty degree (denoted as  $cd$ ) of a proposition is defined as the probability the provenance requester believes the proposition is 'True', that is, the trusted truth value is 'True'. It is used to represent the certainty of the trusted truth value.

$$cd = Pr(TTV). \quad (2)$$

**Definition 3:** the trust degree (denoted as  $td$ ) of a proposition (or the proposition creator) is defined as the degree of belief (subjective probability) that the provenance requester trusts this proposition (or proposition creator).

$$td = Pr(Trusted). \quad (3)$$

### 6.1 Asserted propositions

The trusted truth value of an asserted proposition depends on (i) the truth value assigned by the proposition creator; (ii) the trust degree of the proposition. It is easy



to understand, when the trust degree is 1.0, that the trusted truth value will be the same as the assigned truth value. But when the trust degree is 0.0, what value should the trusted truth value have? Furthermore, when the trust degree is less than 1.0 and greater than 0.0, what is the relation among trusted truth value and assigned truth value as well as trust degree?

First, let us consider the case of trust degree being zero. According to information theory (Shannon 1948, MacKay 2003), 'entropy' is used to measure the uncertainty of information, and the entropy of a variable  $x$  which has  $n$  possible outcomes  $v_1, \dots, v_n$  is defined as follows:

$$H(x) = -\sum_{i=1, \dots, n} p_i \log p_i \quad (4)$$

where,  $p_i$  is the probability for the variable to have outcome  $v_i$  for a given  $n$ , entropy  $H(x)$  is maximum and equal to  $\log n$  when all the  $p_i$  are equal ( $1/n$ ), which is the most uncertain situation; entropy  $H(x)$  is minimum and equal to 0.0 if and only if one  $p_i$  is 1.0 and all others are 0.0, which is the most certain situation. In the case of the variable with only two possible outcomes, the entropy becomes:

$$H(x) = -(p \log p + (1 - p) \log(1 - p)), \quad (5)$$

and the entropy has maximal value when  $p = 0.5$ .

In our context of uncertainty-oriented KP, if the trust degree of an asserted proposition is 0, then no matter what value the truth value assigned by the proposition creator is, there is no information for determining the degree of certainty of the proposition, which is corresponding to the most uncertain situation where the entropy should be maximal. Therefore, the probability of this proposition being 'True' should be 0.5. In Uncertainty-oriented KP, we assign the probability for a proposition to be 'True' as the uncertain truth value of the proposition. Therefore, based on information theory, we assign 0.5 to the trusted truth value of this asserted proposition when trust degree is 0.0, as shown in figure 3(b). As a matter of fact, this situation of asserted proposition can be extended to other types of propositions. When a proposition is distrusted, no matter what type the proposition is, there is no information effective to determine its trusted truth value, so based on information theory, the certain degree of the proposition should be 0.5.

Axiom 6-1:

$$\text{for-all } (a, x) ((\text{type}(x, \text{'KP\_prop'})} \wedge \neg \text{trusted}(x, a)) \\ \rightarrow \text{certainty\_degree}(a, x, 0.5)).$$

Now, consider the general situation when the trust degree is any real value that ranges from 0.0 to 1.0. Recall axiom 1 of Static KP:

$$\text{for-all } (a, x, v) ((\text{type}(x, \text{'asserted\_prop'})} \wedge \text{trusted}(x, a) \wedge \text{assigned\_truth\_value}(x, v)) \\ \rightarrow \text{trusted\_truth\_value}(a, x, v)).$$

which states that an asserted-prop has its trusted truth value as the same as the assigned truth value given by the proposition creator, if the asserted-prop is trusted by the agent making the provenance request. According to this axiom and

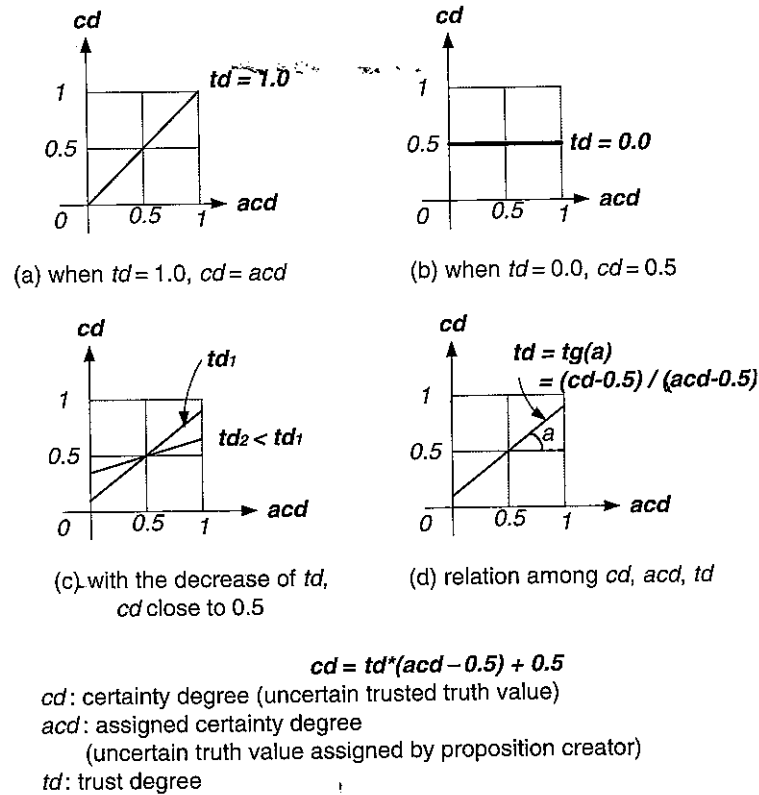


Figure 3. Relation among certainty degree, assigned certainty degree, and trust degree.

probability theory, the probability of the trusted truth value of an asserted proposition being 'True' is calculated by the following formula:

$$\begin{aligned} Pr(TTV) = & Pr(TTV|Trusted, ATV)Pr(Trusted, ATV) \\ & + Pr(TTV|Trusted, \neg ATV)Pr(Trusted, \neg ATV) \\ & + Pr(TTV|\neg Trusted)Pr(\neg Trusted). \end{aligned} \quad (6)$$

As the truth value of a specific proposition assigned by its creator and whether the creator is trusted by provenance requester are independent to each other, according to product rule of probability theory, we have

$$\begin{aligned} Pr(Trusted, ATV) &= Pr(Trusted) \times Pr(ATV) \\ Pr(Trusted, \neg ATV) &= Pr(Trusted) \times Pr(\neg ATV). \end{aligned} \quad (7)$$

Apply (7) and  $Pr(\neg Y) = 1 - Pr(Y)$  to (6):

$$\begin{aligned} Pr(TTV) = & Pr(TTV|Trusted, ATV) \times Pr(ATV) \times Pr(Trusted) \\ & + Pr(TTV|Trusted, \neg ATV) \times Pr(Trusted) \times (1 - Pr(ATV)) \\ & + Pr(TTV|\neg Trusted) \times (1 - Pr(Trusted)). \end{aligned} \quad (8)$$

The conditional probabilities in above formula can be determined as follows. According to axiom 1 of Static KP, when the truth value of a proposition is assigned as 'True', and the proposition (or its creator) is trusted, the trusted truth value is 'True', that is, the probability in which the trusted truth value is 'True' is 1.0, i.e.

$$Pr(TTV|Trusted, ATV) = 1.0. \quad (9)$$

Similarly, when the truth value of a proposition is assigned as 'False', and the proposition (or its creator) is trusted, the trusted truth value is 'False', that is, the probability in which the trusted truth value is 'True' is 0.0, i.e.

$$Pr(TTV|Trusted, \neg ATV) = 0.0. \quad (10)$$

According to information theory and our discussion earlier in this section, when a proposition (or its creator) is distrusted, no matter what the truth value assigned by the proposition creator is, there is no information to determine the trusted truth value, which is corresponding to the most uncertain situation and the 'entropy' has maximal value, so the trusted truth value should be 0.5, i.e.

$$Pr(TTV|\neg Trusted) = 0.5. \quad (11)$$

Applying (9), (10) and (11) and the definitions in (1), (2) and (3) to formula (8), we have

$$ttv = td \times (tv - 0.5) + 0.5. \quad (12)$$

The relation among the certainty degree of a proposition, assigned certainty degree, and trust degree of the proposition, revealed by formula (12), is illustrated in figure 3. When the trust degree is 1.0 (completely trust), the certainty degree is the same as the assigned certainty degree given by the proposition creator (see figure 3(a)); with the decrease of trust degree, the certainty degree is close to 0.5 ('unknown') (see figure 3(c)); when the trust degree is 0.0 (completely distrust), the certainty degree should be 0.5 ('unknown') (see figure 3(b)); if the assigned certainty degree  $acd = 0.5$ , then the certainty degree  $cd = 0.5$ , no matter what the trust degree is (see figure 3(a)-(d)); furthermore, from (12), we could have  $cd - 0.5 = td \times (acd - 0.5)$ , and we know  $td \leq 1$ , so, when the assigned certain degree  $acd > 0.5$ , the certainty degree  $cd \geq 0.5$  and  $(cd - 0.5) \leq (acd - 0.5)$ ; similarly, when  $acd < 0.5$ , then  $cd \leq 0.5$  and  $(0.5 - cd) \leq (0.5 - acd)$ ; in other words, we always have  $|cd - 0.5| \leq |acd - 0.5|$  (see figure 3(d)).

**Theorem 6-1:** The certainty degree of an asserted proposition is dependent on the trust degree of the proposition and the truth value assigned by the proposition creator. The relation among them is:

$$cd = td \times (acd - 0.5) + 0.5. \quad (13)$$

The derivation of formula (12) gives the proof of this theorem.

The same approach used above is applied to set up uncertain KP model for other types of propositions including 'Equivalent', 'Derived', 'AND', 'OR', and 'NEG'. In the following, we only give the conclusions.

## 6.2 Dependent propositions

### 6.2.1 Equivalent propositions

**Theorem 6-2:** The trusted truth value of an equivalent proposition  $x$  is dependent on the trust degree of  $x$  and the trusted truth value of the proposition  $y$  which this equivalent proposition depends on. The relation among them is:

$$cd_x = td_x \times (cd_y - 0.5) + 0.5. \quad (14)$$

The proof of this theorem is similar to Theorem 6-1 and is omitted.

### 6.2.2 Derived propositions

**Theorem 6-3:** The trusted truth value of derived proposition  $x$  is dependent on the trust degree of  $x$  and the truth value assigned by the proposition creator as well as the trusted truth value of proposition  $y$  which  $x$  depends on. The relation among them is:

$$cd_x = td_x \times cd_y \times (acd_x - 0.5) + 0.5. \quad (15)$$

The proof of this theorem is similar to Theorem 6-1 and is omitted.

### 6.2.3 'AND' propositions

Axiom 6-2:

$$\begin{aligned} cd_z &= td_z \times (Pr(x|y) \times cd_y - 0.5) + 0.5 \\ \text{or} \\ cd_z &= td_z \times (Pr(y|x) \times cd_x - 0.5) + 0.5. \end{aligned} \quad (16)$$

### 6.2.4 'OR' propositions

Axiom 6-3:

$$\begin{aligned} cd_z &= td_z \times (cd_x + cd_y - Pr(x|y) \times cd_y - 0.5) + 0.5 \\ \text{or} \\ cd_z &= td_z \times (cd_x + cd_y - Pr(y|x) \times cd_x - 0.5) + 0.5. \end{aligned} \quad (17)$$

### 6.2.5 'NEG' propositions

Axiom 6-4:

$$cd_x = 1 - cd_y. \quad (18)$$

More details about uncertainty-oriented KP can be found in Huang and Fox (2004b).

## 7. Conclusion

As web standards become the dominant means by which information is acquired, stored, searched and delivered throughout the enterprise, the issues of source verification and validity become increasingly important. KP is an approach to

determine the validity and origin of information/knowledge on the web by modelling and maintaining information sources and dependencies as well as trust structures.

We have introduced three levels of KP: Static, Dynamic and Uncertain. Each provides increasingly more realistic forms of support for provenance. To apply KP in practice, knowledge/information creators need to annotate their propositions; knowledge/information users need to define their trust relationships in social networks; and KP reasoner(s) conduct provenance reasoning to deduce the validity. We are now investigating a fourth level of KP where the focus is on how subjective beliefs in trust and validity are acquired from groups.

As stated earlier in the paper: 'we believe the web will always be a morass of uncertain and incomplete information'. The challenge is to create models and processes that enable the creation of islands of certainty within its midst.

### Acknowledgements

This research was supported, in part, by Bell University Laboratory and Novator Systems, Ltd.

### References

- Alexander, J.E. and Tate, M.A., *Web Wisdom: How to Evaluate and Create Information Quality on the Web*, 1999 (Erlbaum: Hillsdale, NJ).
- Allen, J.F. and Ferguson, G., Actions and events in interval temporal logic. *J. Logic Comput.*, 1994, 4(5), 531-579.
- Bartel, M., Boyer, J., Fox, B. LaMacchia, B. and Simon, E., XML-signature syntax and processing. Available online at: [www.w3.org/TR/xmlsig-core/](http://www.w3.org/TR/xmlsig-core/) (accessed 2 May 2003)
- Berners-Lee, T., The semantic web and challenges. Available online at: [www.w3.org/2003/Talks/01-sweb-tbl/](http://www.w3.org/2003/Talks/01-sweb-tbl/) (accessed 26 October 2003).
- Berners-Lee, T., Hendler, J. and Lassila, O., The semantic web. *Scientific American*, May 2001, 84, 35-43.
- Blaze, M., Feigenbaum, J. and Lacy, J., Decentralized trust management, in *Proceedings of IEEE Conference on Security and Privacy*, May, 1996.
- Chu, Y., Trust management for the World Wide Web. Master Thesis, MIT, 1997.
- de Kleer, J., Forbus, K. and McAllester, D., *Truth Maintenance Systems*, IJCAI-89, Tutorial S45, pp. 182-225, 1989.
- Dumbill, E., XML watch: finding friends with XML and RDF. Available online at: [www.106.ibm.com/developerworks/xml/library/x-foaf.html](http://www.106.ibm.com/developerworks/xml/library/x-foaf.html), 2002 (accessed 12 November 2003)
- Fox, M.S. and Huang, J., Knowledge provenance: an approach to modeling and maintaining the evolution and validity of knowledge. EIL Technical Report, University of Toronto, 2003. Available online at: [www.eil.utoronto.ca/km/papers/fox-kp1.pdf](http://www.eil.utoronto.ca/km/papers/fox-kp1.pdf)
- Golbeck, J., Hendler, J. and Parsia, B., *Trust Networks on the Semantic Web*, 2002 (University of Maryland: College Park, MD).
- Goranson, H.T., Huhns, M.N., Nell, J.G., Panetto, H. and Wunram, M., A merged future for knowledge management and enterprise integration. In *Enterprise Inter- and Intra-Organizational Integration*, edited by K. Kosanke, R. Jochem, J.G. Nell and A.O. Bas, pp. 37-50, 2002 (Kluwer Academic: Dordrecht, The Netherlands).
- Gruninger, M. and Fox, M.S., Methodology for the design and evaluation of ontologies. In *Workshop on Basic Ontological Issues in Knowledge Sharing*, IJCAI-95, Montreal, 1995.

- Huang, J. and Fox, M.S., Dynamic knowledge provenance, in *Proceedings of Business Agents and Semantic Web Workshop*, London, Canada, 2004a, pp. 11–20.
- Huang, J. and Fox, M.S., Uncertainty in knowledge provenance. In *The Semantic Web: Research and Applications, Lecture Notes in Computer Science 3053*, edited by C. Bussler, J. Davies, D. Fensel and R. Studer, pp. 372–387, 2004b (Springer: Berlin).
- Khare, R. and Rifkin, A., Weaving and web of trust. *World Wide Web Journal*, 2(3), 1997, 77–112.
- MacKay, D.J.C., *Information Theory, Inference, and Learning Algorithm*, 2003 (Cambridge University Press: Cambridge).
- McGuinness, D.L. and Pinheiro da Silva, P., Infrastructure for web explanations. In *Proceedings of the International Semantic Web Conference*, Florida, 2003, pp. 113–129.
- Oliver, K., Evaluating the Quality of Internet Information. Virginia Tech, 1997.
- Richardson, M., Agrawal, R. and Domingos, P., Trust management for the semantic web, in *Proceedings of the International Semantic Web Conference*, Florida, 2003, pp. 351–368.
- Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27, 1948, 379–423, 623–656.
- Simon, E., Madsen, P., Adams, C., An introduction to XML digital signatures, Available online at: [www.xml.com/pub/a/2001/08/08/xmldsig.html](http://www.xml.com/pub/a/2001/08/08/xmldsig.html), 2001 (accessed 16 January 2003).
- Tan, K., Building your appropriate certificate-based trust mechanism for secure communications. Available online at: [www.rainbow.com/library/8/BuildingYourAppCert\\_based.pdf](http://www.rainbow.com/library/8/BuildingYourAppCert_based.pdf), 2002 (accessed 16 January 2003).
- Watts, D.J., *Small Worlds: The Dynamics of Networks Between Order and Randomness*, 1999 (Princeton University Press: Princeton, NJ).
- Yu, B. and Singh, M.P., A social mechanism of reputation management in electronic communities, in *Proceedings of Fourth International Workshop on Cooperative Information Agents*, 2000, pp. 154–165.
- Zimmermann, P.R., *The Official PGP User's Guide*, 1995 (MIT Press: Cambridge, MA).