Knowledge-based Financial Statement Fraud Detection System: Based on an Ontology and a Decision Tree†

Xiao-Bo Tang*, Guang-Chao Liu**, Jing Yang***, Wei Wei****

*/**/*** Wuhan University, Center for the Studies of Information Resources, Wuhan, Hubei, China,
**** Research Institute of Big Data, Zhongnan University of Economics and Law, Wuhan, Hubei, China
* <xbtang2010@126.com>, **<chao1635@126.com>,
*** <591486903@qq.com>, ****<503175355@qq.com>

Xiao-Bo Tang is a professor of information science at the Information Management School of Wuhan University in China. He holds a PhD in management science and engineering from Wuhan University and a master's degree in computer science and technology from Wuhan University of Hydraulic and Electrical Engineering. He is the director of the Information Systems Research Center in Wuhan University, member of the Association for Information Systems, the director of the China Branch of the Association for Information Systems, and the executive director of the Information Society of Hubei Province. His main research interests include knowledge organization and intelligence analysis.

Guang-Chao Liu is a doctoral student at the Information Management School of Wuhan University since September 2015. He completed his MS in financial management from the Lubin School of Business, Pace University. He is currently working on text mining in the financial domain at the semantic level. His research interests also include sentiment analysis and knowledge organization.

Jing Yang is a postgraduate student at the Information Management School of Wuhan University since September 2016. She got her bachelor's degree in management from Business School, Jianghan University. She is currently working on text mining with machine learning and deep learning. Her research interests also include knowledge organization and intelligence service.

Wei Wei is a lecturer of data science at Big Data Institute of Zhongnan University of Economics and Law in China. She holds a PhD in management science and engineering from Wuhan University and attained her master's and bachelor's degrees in management information system. Her research interests lie in the knowledge organization and intelligence service.

Tang, Xiao-Bo, Guang-Chao Liu, Jing Yang and Wei Wei. 2018. "Knowledge-based Financial Statement Fraud Detection System: Based on an Ontology and a Decision Tree." *Knowledge Organization* 45(3): 205-219. 45 references. DOI:10.5771/0943-7444-2018-3-205.

Abstract: Financial statement fraud has seriously affected investors' confidence in the stock market and economic stability. Several serious financial statement fraud events have caused huge economic losses. Intelligent financial statement fraud detection has thus been the topic of recent studies. In this paper, we developed a knowledge-based financial statement fraud detection system based on a financial statement detection ontology and detection rules extracted from a C4.5 decision tree algorithm. Through discovering the patterns of financial statement fraud activity, we defined the scope of our financial statement domain ontology. By utilizing SWRL rules and the Pellet inference engine in domain ontology, we detected financial statement fraud activities and discovered implicit knowledge. This system can be used to support investors' decision-making and provide early warning to regulators.

Received: 21 December 2017; Revised: 25 February 2018; Accepted: 20 March 2018

Keywords: financial statement fraud detection, ontology, decision tree

† This study is part of the National Natural Sciences Foundation of China "Research on Intelligence Consulting Services Based on the Semantic Analysis of Text and Web" (No. 71673209).

1.0 Introduction

Financial statement fraud happens when corporations intentionally prepare financial statements that include misstated or misrepresented material to mislead stock market investors and regulators (Rezaee 2005). According to Hajek and Henriques (2017), the common types of financial statement fraud include omissions in financial records, falsification or manipulation of revenue, income, assets, expenses and other financial variables, and misrepresentation of management discussions and analysis.

Financial statement fraud seriously affects investors and regulators. It causes huge losses in the economy and the stock market and destroys the general public's confidence in the business environment. In the past few years, several firms have been involved in financial statement fraud activity, which led to economic turmoil. For example, Enron and other firms perpetrated financial fraud, which enormously affected the world economy and stock market (Dong 2014). According to Abbasi (2012), in the ten largest bankruptcies in United States history, four companies were involved in major financial fraud. Beasley et al. (2010) showed that in firms that commit fraud, 28% were bankrupted in two years, and 47% were delisted from the stock exchange. Therefore, financial statement fraud has attracted much concern from investors and regulators.

Detecting financial statement fraud requires expert knowledge and experience. According to West et al. (2016), traditional methods for financial statement fraud detection are time consuming, expensive, and inaccurate. Dyck et al. (2010) illustrate that from 1996 to 2004, most fraud activities were not detected by regulators or auditors. In this paper, we propose a knowledge-based system for financial statement fraud detection based on an ontology, SWRL, and a decision tree algorithm. In this research, we build an ontology of financial statements and use a decision-tree algorithm to find financial statement fraud patterns and transform those patterns to SWRL rules that can be used in a knowledge-based system. The remainder of this paper is organized as follows: section 2 reviews previous research on financial statement fraud detection, ontology, the decision tree's rules extraction, OWL, and SWRL. Section 3 presents the research model for this study. Section 4 describes the experiment's material and method for financial statement fraud detection, including datasets, ontology construction, SWRL rules and inference model. Section 5 presents the experiment and discusses the test results. Section 6 concludes the paper and presents future work.

2.0 Literature review

2.1 Intelligent financial statement fraud detection

Intelligent financial statement fraud detection aims at discovering the patterns of financial statement fraud to provide early warning to regulators and support investors' decision-making processes by using artificial intelligence methods. In the detection of financial statement fraud, internal data and external data are major source materials. According to Abbasi (2012), internal data include "auditor-client relationships, personal and behavioral characteristics, internal control overrides and so on." Cecchini et al. (2010) point out that those internal data were not open to investors and other stock market participants. Access to internal data is difficult and time consuming to achieve (Abbasi et al. 2012). Publicly listed firms' financial statements are compulsorily disclosed by regulators and laws. Those external data are easily accessible and highly reliable. In this study, we limit our discussion to the previous research that used publicly available data and machine-learnable methods. A list of previous studies is presented in Table 1, including authors' names, feature set, data set, method, and accuracy.

Financial variables are important indicators in financial statement fraud detection. Financial variables from financial statements can reflect companies' financial performances in many aspects (Hajek and Henriques 2017). Auditing financial statements can uncover whether firms are involved in a crisis (Ravisankar et al. 2011). The pressure of involvement in a crisis may prompt managers to improve companies' financial performance by using illegal methods (Bell et al. 1991). Companies' financial performances can be depicted by several financial variables, such as current assets, current liabilities, total assets, and other ratios (Summers et al. 1998).

Machine-learning methods and ontology-based methods are used in financial statement fraud detection. Kanellopoulos et al. (2007) proposed a web service framework for publicly traded Greek manufacturing firms' financial statement fraud detection based on an ontology. They proposed a software structure that was constructed by a semantic web layer and an internal layer. The semantic web layer contained ontologies about firms and auditors and an interface that was available to users. The internal layer contained control and reasoning components. This service system can help users find fraud activity by using an ontology along with a reasoning engine. Machine-learning methods are commonly used in financial statement fraud detection. Table 1 presents the previous studies of financial statement fraud detection by using machine learning. "Fraud" firms indicate those who have committed financial fraud, and "non-fraud" firms indicate those who have not.

Author	Feature Set	Data Set	Method accuracy
Kirkos et al. (2007)	10	38 fraud Greek firms; 38 non-fraud Greek firms	Bayesian Belief Network-90.3%, Multilayer Perceptron-80%, ID3-73.6%
Ravisankar et al. (2011)	18	101 non-fraud Chinese firms; 101 fraud Chinese firms	Probabilistic Neural Network-98.1%, Genetic Programming -94.1%, Group Method Data Handling-93%, Multilayer Perceptron -78.8%, Support Vector Machine-73.4%
Pai et al. (2011)	18	25 fraud Taiwanese firms; 50 non-fraud Taiwanese firms	Support Vector Machine-92%, C4.5-84%, Radial Basis Function Neural Network-82.7%, Multilayer Perceptron -82.7%
Abbasi et al. (2012)	12	815 yearly and quarterly fraud instances from U.S.; 8191 non-fraud yearly and quarterly instances from U.S.	Support Vector Machine-Linear-90.4%, Support Vector Machine-Polynomial-86.5%, Support Vector Machine-Radial Basis Function-89.4%, Naïve Bayes-85.1%, Bayesian Networks-89.3%, J48-83.9%, Naïve Bayes Tree-88.7%, ADTree-89.6%, Random Forest-85.7%, REPTree-88.8%, Nearest Neighbor-86.5%, JRip-87%, Logistic Regression-87.5%, Neural Networks-86.5%
Song et al. (2014)	23	10 fraud Chinese firms; 440 non-fraud firms	Voting-88.9%, Support Vector Machine-85.5%, Multi-layer Perceptron-85.1%, C5.0-78.6%
Chen et al. (2014)	8	66 fraud Taiwanese firms; 66 non-fraud Taiwanese firms	C5.0-85.7%, Logistic Regression-81%, Support Vector Machine-72%
Liu et al. (2015)	8	138 fraud Chinese firms; 160 non-fraud Chinese firms	Random Forests -88%, Support Vector Machine-80.18%, CART-66.43%, k-NN (60.11), Logistic Regression-42.91%
Omar et al. (2017)	10	15 fraud Malaysia firms; 95 non-fraud Malaysia firms	ANN-94.87%
Petr Hajek and Roberto Henriques (2017)	14	311 fraud U.S. firms; 311 non-fraud U.S. firms	Logistic Regression-77.31%, Naïve Bayes-61%, Bayesian Belief Network-90.05%, Decision Table/Naïve Bayes Hybrid Classifier-90.09%, Support Vector Machine-80.5%, JRIP-86.95%, C4.5-86.6%, CART-87.09%, Logistic Model Trees-86.26%, Multilayer Perceptron-85.13%, Voted Perceptron-49.59%, Bagging-87.84%, Random Forests-88.89%, AdaboostM1-80.5%

Table1. Previous studies of financial statement fraud detection using machine learning.

In the studies presented in Table 1, the most commonly used classification methods were support vector machines, logistic regression, decision trees, and neural networks. Hajek and Henriques (2017) employed fourteen classification methods for fraud classification. Decision table/naïve Bayes hybrid classifier achieved the best performance, which was 90.09%, and Bayesian Belief Network achieved a similar performance of 90.05%. In that research, the highest accuracy rate was achieved by Ravisankar et al. (2011), which was 98.1%, by using Probabilistic Neural Network. Most studies employed annual financial statements as experiment data, and four studies in the list used a pairing method in which the number of fraud- and non-fraud-committing firms, size, industry, and corresponding year were matched.

2.2 Ontology

An ontology is "a formal, explicit specification of a shared conceptualization" (Gruber 1993). Specifically, ontology formally describes concepts in a domain and those concepts' attributes (Noy et al. 2000). By abstracting the concepts and terminology of a specific domain, the ontology forms the shared concepts of a domain and constructs a domain's conceptual model. The elements of an ontology are classes, instances, and properties (Campos et al. 2009). A class is an abstract description of a set of collections with the same characteristics. An ontology is usually composed of multiple classes and therefore forms a concept set. "Properties" are descriptions of relationships between ontology classes. An "instance" is the most specific object in the class. If an individual is subordinate to a class, it means that the individual is an instance of that class. Ontology is widely applied in knowledge organization domains including medical, financial, and other domains.

Some works have been written on ontologies for the financial domain. The Financial Fraud Prevention Oriented Information Resources using Ontology (FF POIROT) project provided multilingual semantic web services in the financial forensics domain (Kingston et al. 2004). A computable and shareable knowledge domain of the financial fraud area of European law was constructed by this project to help law enforcement departments solve financial fraud problems by using a novel method. Zhao et al. (2004) pointed out that FFPOIROT was developed based on the DOGMA ontology paradigm. Shue et al. (2009) developed an ontology-based expert system for financial statement analysis to predict future conditions and

performance of firms. They employed an ontology to represent the domain knowledge of financial statements and used decision rules to do inference processes. The Financial Industry Business Ontology (FIBO) program, which is being developed by the Enterprise Data Management Committee (EDM), has gained wider recognition in recent years (Bennett 2013). FIBO aims to semantically model all financial terms and financial relationships and provides machine-readable standardized information so as to create highly automated conditions for finance information collection, processing, and anonymous sharing.

2.3 Rules extraction from a decision tree algorithm

Decision tree algorithms are among the most commonly used data mining techniques with a fast learning speed, and they produce classification rules that are easy to understand (Han 2011). The decision tree is a tree structure that starts with a single root node. The leaf nodes of the tree store some class label values, indicating a possible classification result. A path from the root node to the leaf node forms a classification rule, and a decision tree can be easily transformed into several classification rules. The ID3 algorithm is based on information theory. In ID3, the best splitting attributes are chosen based on the highest information gain. The information gain is measured by entropy. The entropy is defined as:

$$Entropy(S) = -\sum_{i=1}^{k} Pi \log_2 Pi \qquad (1)$$

Where S is the sample dataset, Pi is the proportion of dataset S belonging to class i.

Gain (S, A) is the information gain of sample dataset S.

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{k} \frac{|Si|}{|S|} Entropy(Si) \qquad (2)$$

Where Si is the subset of S, $|Si|$ is the number of elements of Si. $|S|$ is the number of elements of S. A is the attribute.

Based on the deficiency of the ID3 algorithm, Quinlan (1993) proposed a modified decision tree algorithm: C4.5. The C4.5 algorithm inherits the advantages of the ID3 algorithm and makes some improvements to the ID3 algorithm:

1) Processes continuous data and discrete data;
2) Processes data with missing values;
3) Uses information gain ratio as the feature selection criteria.

In C4.5, the best splitting attributes are chosen based on the gain ratio. The gain ratio is defined as equations (3) and (4):

$$Splitinfo(S,A) = -\sum_{i=1}^{k} \frac{|Si|}{|S|} \times log_2\left(\frac{|Si|}{|S|}\right) \qquad (3)$$

$$GainRatio(A) = \frac{Gain(S,A)}{Splitinfo(S,A)} \qquad (4)$$

Quinlan (1987) pointed out that a decision tree, as a classification algorithm, could acquire knowledge and extract production rules from the generated tree structure. The canonical format of production rule is:

IF P THEN Q

These rules can be obtained along the path from the root node of the decision tree to the leaf nodes, where each attribute-value pair along a given path constitutes the antecedent of the rule (the "IF" part). The leaf node gives the category of the predicted value and constitutes the consequent of the rule (the "THEN" part). The IF-THEN rules can easily predict unknown samples.

In financial statement fraud detection, we process continuous numerical data, and data with missing values may exist. Based on those two points, we chose the C4.5 algorithm to generate rules.

2.4 OWL and SWRL

OWL is an ontology representation language, which is based on description logic (McGuinness and Harmelen 2006). OWL is one of the core languages of the semantic web for knowledge representation (Padmavathi and Krishnamurthy 2017). It inherits the basic fact statement of RDF and the class and attribute structure of RDF schema (Horrocks et al. 2003). It complements the deficiency of RDF/RDFS that cannot describe relationships well. The OWL language can clearly describe the concept of knowledge and the relationships between concepts.

SWRL (Semantic Web Rule Language) is proposed to improve OWL's inference ability. SWRL is based on the combination of OWL DL and OWL Lite (Horrocks et al. 2004). SWRL can describe rules and infer new knowledge from OWL individuals (O'Connor et al. 2012). The SWRL format is "antecedent→ consequent," where antecedent and consequent are the conjunction of atoms in the ontology (Pan et al. 2005). For example, a person has a female sibling, and then the person and the female sibling are sisters. SWRL can express this inference as follows:

Person (? x) ^ hasSibling (? x, ? y) ^ Female (? y)→ hasSister(? x, ? y)

3.0 Model framework

This study aims at developing a model for the expression and inference of the patterns of financial statement fraud in order to provide early warning to regulators and support investors' decision-making processes. The model framework is presented in Figure 1.

1) The knowledge-based financial statement fraud detection system contains a fraud detection ontology, SRWL rules, and Pellet inference engine. The preprocessed data from financial statements include financial variables that can be used in a C4.5 decision tree algorithm for financial statement fraud detection rules extraction and fraud detection ontology construction. The C4.5 decision tree can be used to discover financial statement fraud patterns from datasets. The inference rules are extracted from the C4.5 decision tree algorithm, which can generate decision tree rules (Quinlan 1987).
2) The fraud ontology contains firms and financial variables that are chosen based on a selection of financial variables. Besides class and property definitions, instances are added into the ontology. The production rules generated from the decision tree can be described by SWRL. SWRL rules are constructed by classes, properties, and instances from the existing ontology. The fraud detection ontology and SWRL rules normalize the knowledge of the financial statements and the knowledge of fraud activities and create a knowledge base for financial statement fraud detection.
3) The inference engine is used to identify which instances contain fraud activity or not. The inference engine can convert OWL and SWRL into a format that can be used in the inference process. The inference results can be written into OWL and update the domain ontology. This knowledge-based system can provide users early warning about whether a firm has potentially committed fraud.

4.0 Method

4.1 Data collection

The instances of fraudulent financial statements were identified from Accounting and Auditing Enforcement Releases (AAERs), released by the U. S. Securities and Exchange Commission (SEC). The firms that violate federal or SEC rules are disclosed by an AAER, and the SEC takes actions against these firms. Dechow et al. (2011) pointed out that AAERs are highly authoritative, because the SEC would take enforcement action only when the firm showed strong evidence of fraudulent activity. The SEC alleges that firms are involved in fraudulent activity based on Rule17 (a) from the Security Exchange Act of 1933, Rules 13(b)(5), 13b2-1, and Rule 10(b)-5 from the 1934 Securities Exchange Act (Cecchini et al. 2010). In this research, we identified 130 firms involved in AAER reporting during the period of 1998-2016. A set of 130 fraudulent annual reports were employed in our research. Annual reports are the ideal sample for fraud detection, because they contain financial information that can reflect a firm's financial status. In order to match the sample of firms committing fraud, we identified firms of similar size that did not commit fraud within the same year and industry. Our dataset

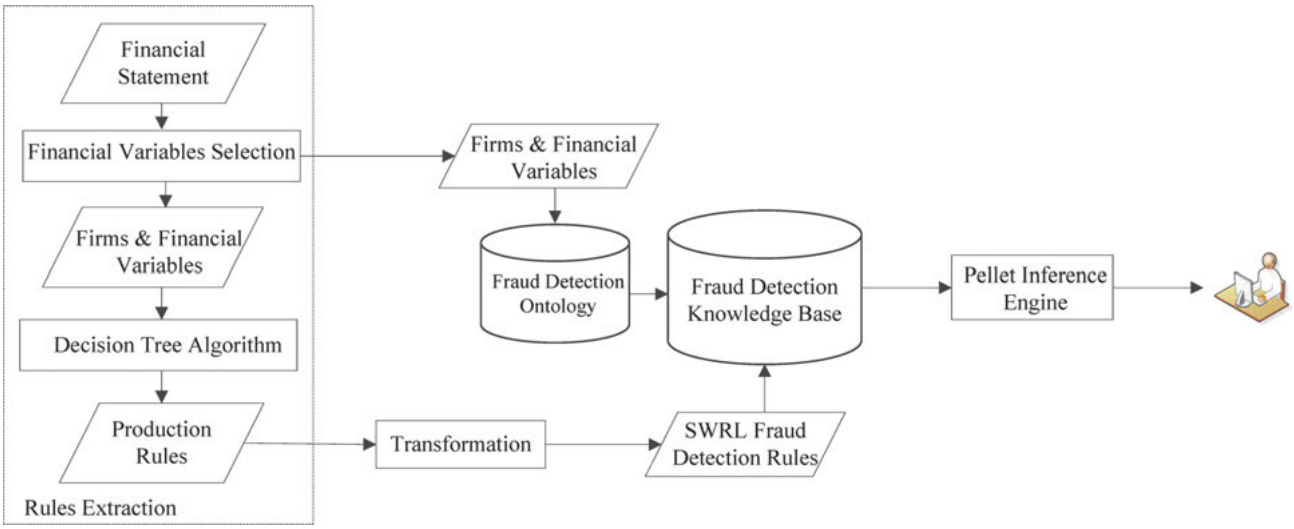

Figure 1. Framework for financial statement fraud detection.

contained 260 firms (130 fraud-committing firms and 130 non-fraud-committing firms). In this dataset, 200 firms were employed in the generation of fraud detection rules from C4.5, and sixty firms were used in ontology construction.

Financial variables are the critical indicator in financial statement fraud detection. The financial variables chosen in financial statements can reflect all aspects of firms' financial status to detect different types of financial statement fraud. Previous studies show strong evidence that financial variables are useful for financial statement fraud detection. The financial variables that we employed in this research are presented in Table 2 and are based on previous studies (Kotsiantis et al. 2006; Kirkos et al. 2007; Ravisankar et al. 2011; Li et al. 2014; Liu et al. 2015; Hajek and Henriques 2017). Those financial variables can be classified into five categories: firm size, profitability variables, operational variables, structure variables, and activity variables.

Firm size	total assets
	revenue
Profitability variables	net income
	net income / revenue
	ROE(return on equity)
	ROA(return on assets)
	profit margin
	gross margin
	EBITDA margin
	EBIT margin
	cash flow/operating revenue
Operational variables	net assets turnover
	stock turnover
Structure variables	current ratio
	liquidity ratio
Activity variables	revenue/total assets
	cash flow/total assets
	cash flow

Table2. Financial variables used for variable selection.

4.2 Variable selection

In order to reduce data dimensionality and improve accuracy, we employed feature selection on the dataset. The results of variable selection will be used in domain ontology construction and inference rule extraction. Feature selection focuses on choosing a subset of variables from all variables and can minimize irrelevant variables and improve the performance of predictors (Guyon and Elisseeff 2003). By applying feature selection, we can improve the generalizability, comprehensibility, and computational efficiency of the model.

In this research, we employed an extremely randomized tree algorithm, also known as an extra-tree algorithm, as a feature selection method (Geurts et al. 2006). Extremely randomized trees feature selection is an ensemble method that integrates feature selection as a part of the training process based on tree models. Opitz (1999) illustrated that the ensemble method combined several separately trained algorithms, which can improve the accuracy of generalization.

Different from traditional feature selection, ensemble feature selection selects multiple attribute subsets to generate individual learners in order to improve the integration of individual difference (Liu 2007). Extremely randomized trees algorithms develop random forest algorithms. Differing from random forest using bootstrap replica to generate the learning sample, extra-tree employs a whole learning sample to generate the tree (Geurts et al. 2006). The splitting points of extra-tree algorithms are generated by randomizing the selection of candidate variables. The extra-tree method uses the difference of variable-importance to measure the prediction strength of each variable. The variable-importance measure is constructed by out-of-bag samples from random forests (Hastie et al. 2001). Instead of using out-of-bag samples, extra-tree algorithms employ whole learning samples to construct the difference of variable-importance. This can reduce the variance and improve the generalizability of the model. Figure 2 shows the variable-importance in our research. We chose the top five variables that are equal to or greater than 0.07 in the ontology construction and rule generation. Those variables are profit margin, net assets turnover, ROE, cash flow/operating revenue, and revenue.

4.3 Domain ontology construction

The construction of the financial statement fraud detection ontology used financial variables from financial statements to detect firms' fraudulent activities. The process of ontology construction includes identifying the ontology domain and important terminologies and defining class, class hierarchical structure, and attributes. In this study, we constructed this ontology using Protégé 5.2. Protégé is a Java-based tool that integrates ontology editing and knowledge-base editing developed by Stanford University. It provides users with a graphical interface, interactive ontology design, and development environment. Protégé supports class, class multiple inheritance, class attributes, and examples of knowledge representation elements and can define a variety of knowledge rules. Protégé, as open source software, provides a large number of plug-ins and supports XML, RDF/RDFS, OIL, DAML+OIL, and

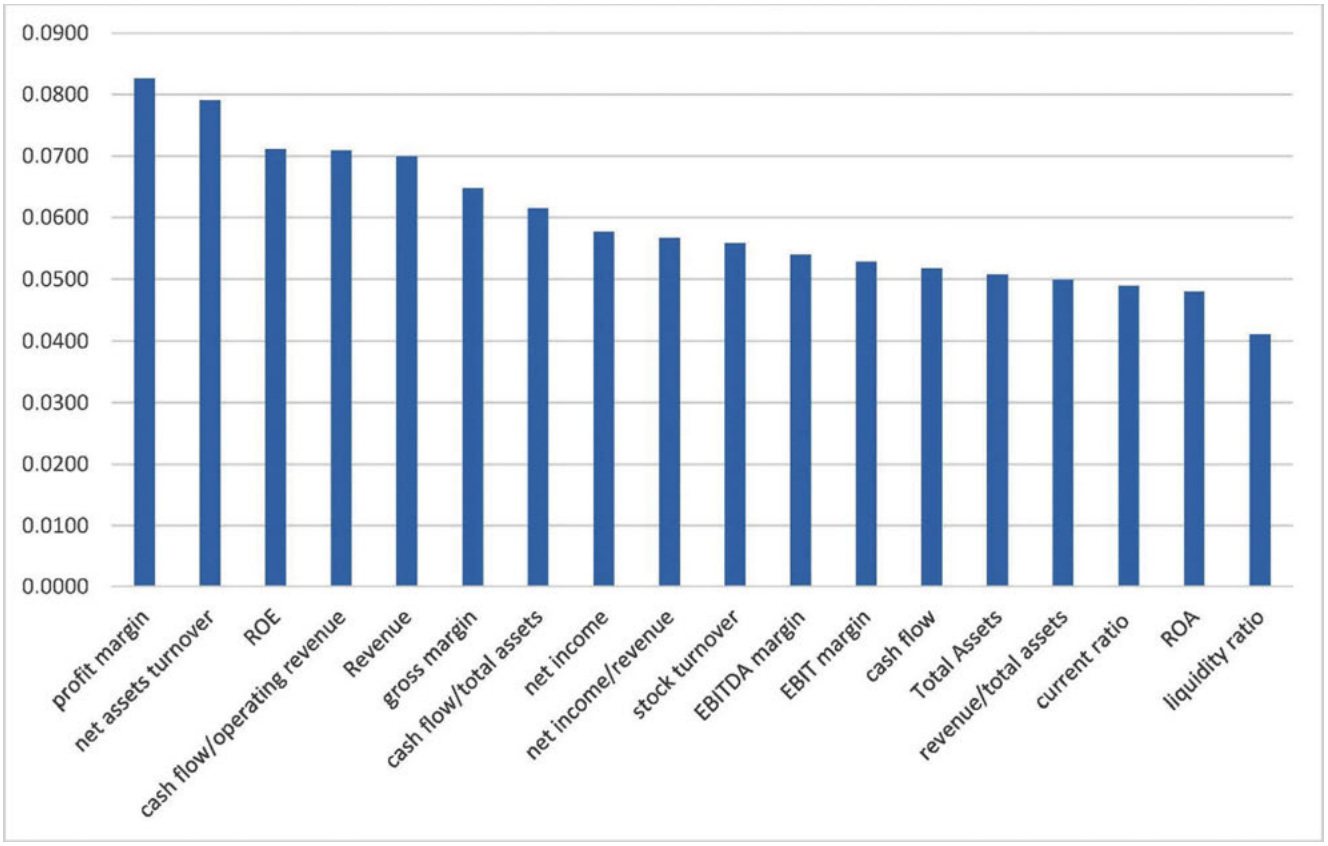

Figure 2. Variable selection based on variable-importance.

OWL. In this study, we used VOWL, which is a Protégé plug-in, to visualize the financial statement fraud detection ontology. Figure 3 shows the model of fraud detection ontology. In Figure 3, class, subclass, object properties, and data properties are shown.

4.3.1 Class definition

Classes are used to describe abstract entity objects. Classes are inherited and organized in the form of hierarchies. The top classes represent the most abstract conceptualizations of entities. Subclasses inherit the abstract properties of their parent classes and represent more specific entity conceptualizations. The fraud detection ontology contains three classes: financial statement, firm, and financial variables. In the "firm" class, firms were classified by their four-digit Standard Industrial Classification Code (SIC). All publicly listed companies in the U.S. stock market have SIC codes. In this ontology, the industries' SIC codes were used in ontology construction. The firm class has thirty-one subclasses of industries based on SIC codes. The financial variables class contains five subclasses: ROE, revenue, profit margin, net assets turnover, and cash flow/operating revenue. Figure 4 shows the fraud detection ontology.

4.3.2 Property definition

In ontology construction, properties are used to describe the common features of a class or the proprietary features of some individual instances. The OWL ontology contains two important properties: object properties and datatype properties. The object properties describe the relationship between two classes. Datatype properties represent a class's own attributes. In financial statement fraud detection ontology construction, isPartof, hasFinancialstatement, and hasFinancialvariable are defined as object properties. In datatype properties, hasValue and hasFraudActivity are defined.

4.3.3 Adding instances and consistency test

Based on the experiment's requirements, we added instances of sixty firms into the firm class and 300 financial variables instances into the financial variables class respectively. Figure 5 shows the instances in Protégé. In order to guarantee that no contradictory knowledge exists in the ontology model, we used the Pellet reasoner to conduct a consistency test. Figure 6 shows the results of the consistency test.

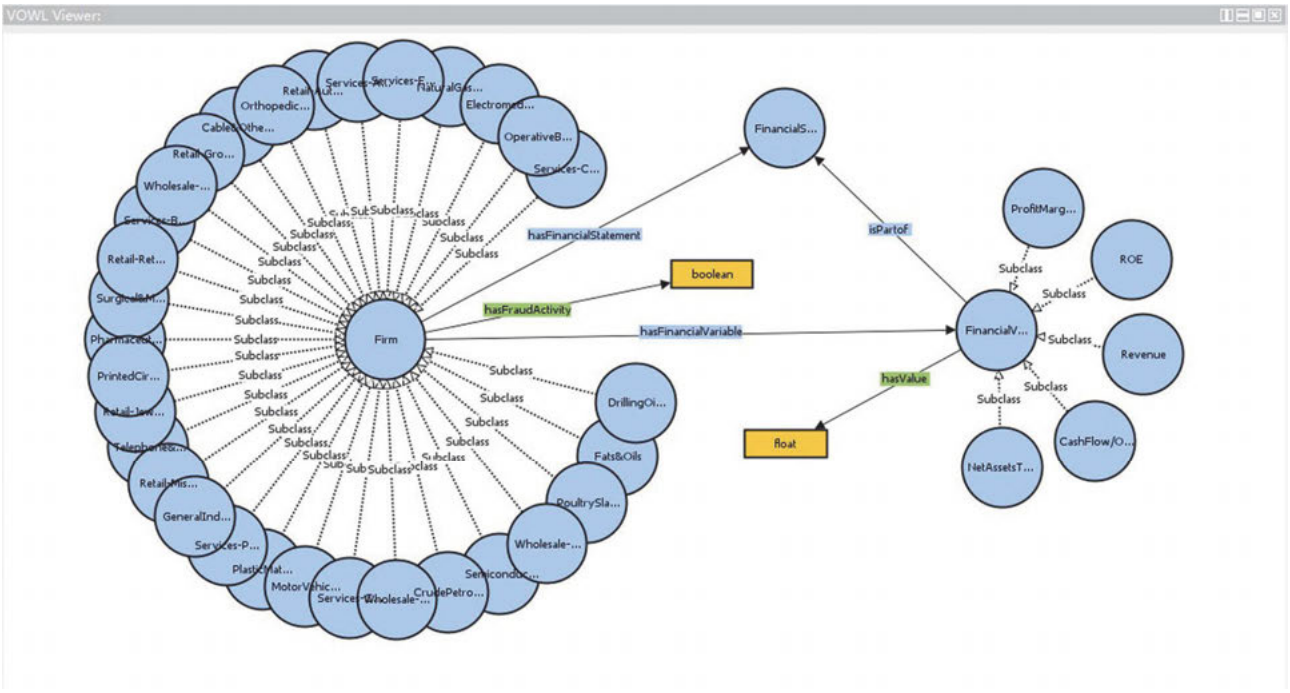

Figure 3. The model of financial statement fraud detection ontology.

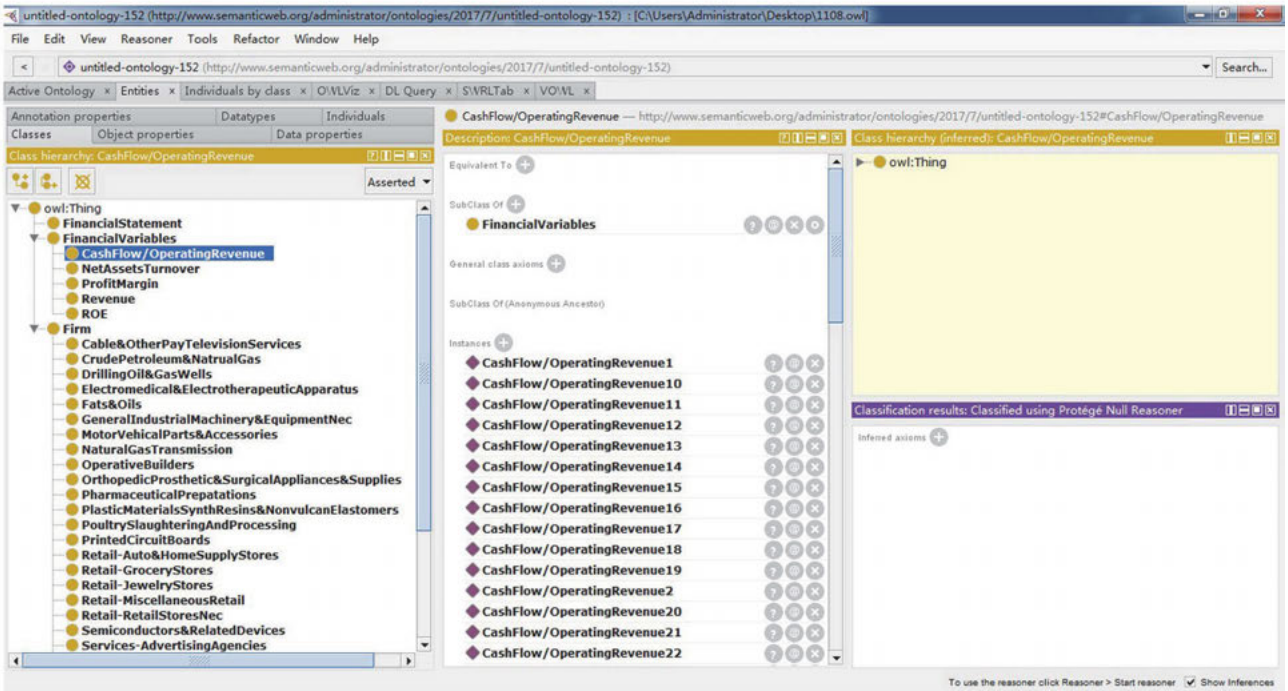

Figure 4. Financial statement fraud detection ontology developed by Protégé.

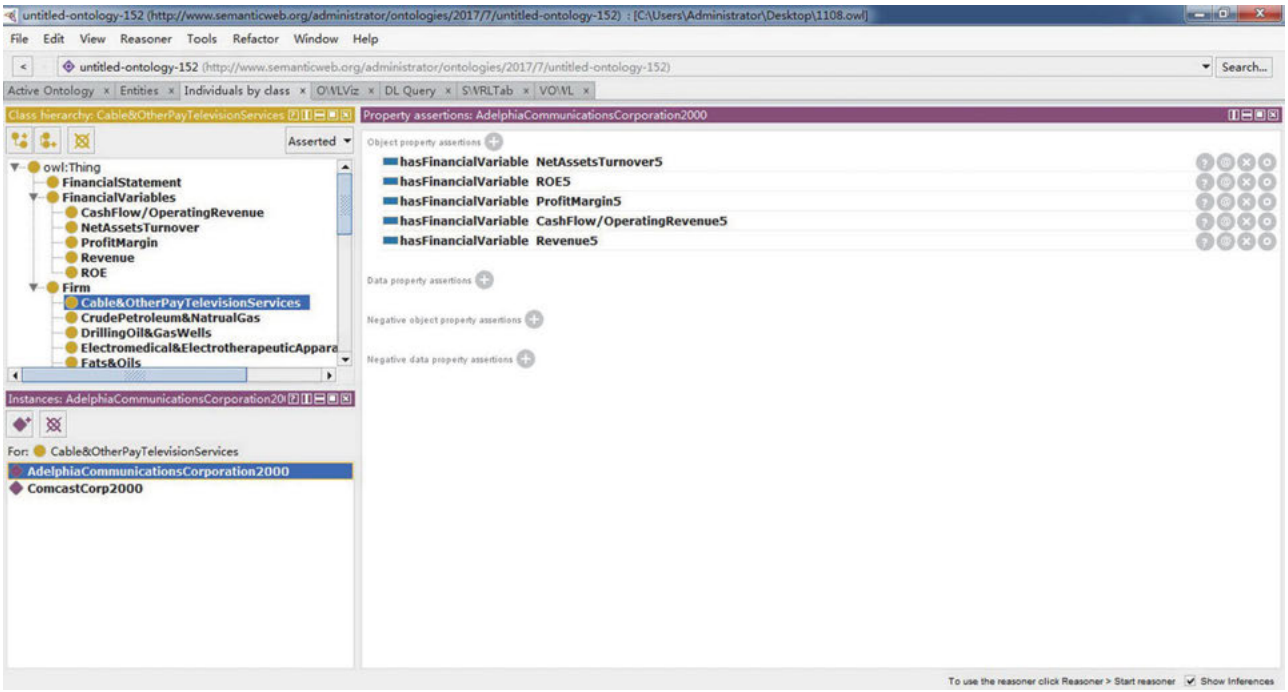

Figure 5. Instances in the financial statement fraud detection ontology.

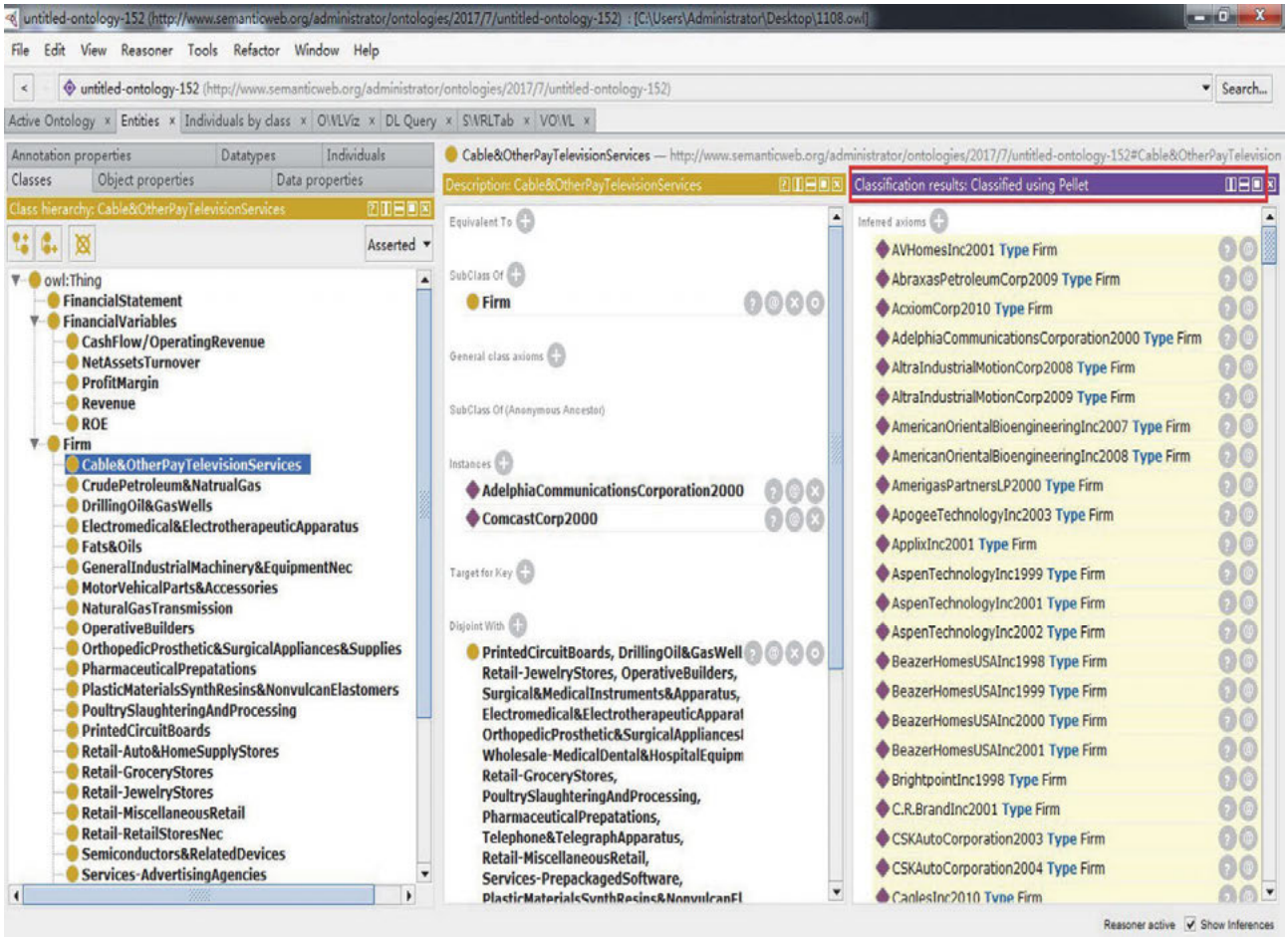

Figure 6. Consistency test in Protégé.

4.4 Rule extraction from C4.5

A decision tree is an efficient and powerful classification algorithm that is popular for classifying patterns of a dataset. The C4.5 is a modified algorithm of a decision tree from which decision rules can be extracted. In this study, we used 200 firms' data to generate fraud detection rules from C4.5. In order to lower the bias and variance of the model and to choose the best classifier for rule extraction, we used k-fold cross-validation to estimate the accuracy of the classifier. In k-fold cross-validation, also called rotation estimation, the dataset D is randomly split into k mutually exclusive subsets (the folds: D_1, D_2, ..., D_k) of approximately equal size. In k times experiments, one subset Dt (t ∈ { 1, 2, ..., k}) was chosen as a tested dataset of each experiment; the rest of the k-1 datasets are trained datasets. The model was trained and tested k times.

To evaluate the performance of the classifier, a ten-fold cross-validation method was employed. Previous studies illustrate that ten-fold cross-validation might be an optimal method to minimize bias and variance (Kohavi 1995). In ten-fold cross-validation, the dataset is equally split into ten folds randomly, and each subset is employed once as a test dataset to test the performance of classifier generated by the remaining nine folds. Based on the ten-fold cross-validation, a decision tree of which accuracy is 81% and f-measure is 80%, was selected for rule extraction.

Based on the decision tree and the path from the root node to the leaf node in the decision tree, we extracted nine rules:

1) If ROE <= -0.0433, then the firm shows fraudulent activity;
2) If ROE > -0.0433 and Revenue <= 12650.0 and Profit Margin <= 0.3625 and Cash Flow/Operating Revenue, then the firm shows fraudulent activity;
3) If ROE > -0.0433 and 12650 < Revenue <= 2630000 and Profit Margin <= 0.3625 and Cash Flow/Operating Revenue <= 0.3505 and Net Assets Turnover <= 1.65, then the firm shows no fraudulent activity;
4) If ROE > -0.0433 and 12650 < Revenue <= 2630000 and Profit Margin <= 0.3625 and Cash Flow/Operating Revenue <= 0.3505 and Net Assets Turnover > 1.65, then the firm shows fraudulent activity;
5) If ROE > -0.0433 and Revenue <= 2630000 and Profit Margin <= 0.3625 and Cash Flow/Operating Revenue > 0.3505, then the firm shows no fraudulent activity;
6) If ROE > -0.0433 and Revenue <= 2630000 and Profit Margin > 0.3625, then the firm shows fraudulent activity;
7) If ROE > -0.0433 and 2630000 < Revenue <= 7465000 and Net Assets Turnover <= 3.695, then the firm shows no fraudulent activity;
8) If ROE > -0.0433 and 2630000 < Revenue <= 7465000 and Net Assets Turnover > 3.695, then the firm shows fraudulent activity;
9) If ROE > -0.0433 and Revenue > 7465000, then the firm shows fraudulent activity.

4.5 Inference model and SWRL

The purpose of ontology reasoning is to obtain implicit knowledge from explicit knowledge. The inference engine has two main functions. The first function is to check the consistency and integrity of the ontology in the process of ontology construction and to ensure that there is no conflict between classes and instances. The second function is to obtain implicit knowledge from the ontology through rules.

An ontology inference engine is based on description logic. Pellet, Racer, and FaCT++ are typical ontology inference engines (Abburu 2012). Those inference engines have the advantages of convenient usability and high reasoning efficiency. In this study, we used Pellet as the ontology inference engine. Pellet is an open-source description logic reasoner based on tableaux algorithms (Sirin et al. 2007). Pellet can support datatype reasoning, SWRL rules, and ontology consistency and integrity checks. Figure 7 shows the workflow of the Pellet inference engine. First, the inference engine reads the OWL file and then converts it into a tuples format with a parser. Second, the inference engine performs species verification and ontology repair and then loads the ontology file into the inference engine. Tbox is used to store class axioms during loading of ontology files, and Abox is used to store individuals. Third, the Tableau reasoner performs reasoning based on Tbox and Abox.

In this study, we used SWRL format to describe rules for the inference engine to financial statement fraud detection. SWRL can provide a semantic complement for an OWL ontology, so as to realize the semantics that the OWL ontology cannot describe.

In the editing of SWRL, rules, classes, instances, and properties can be used directly. Based on the rules extracted from the C4.5 decision tree, we transformed nine production rules into nine SWRL rules for fraud detection. Those rules are described as the following:

1) Firm(?f)^ROE(?r)^hasFinancialVariable(?f,?r)^has Value(?r,?v1)^swrlb:lessThanOrEual(?v1,-0.0433)-> hasFraudActivity(?f,true);

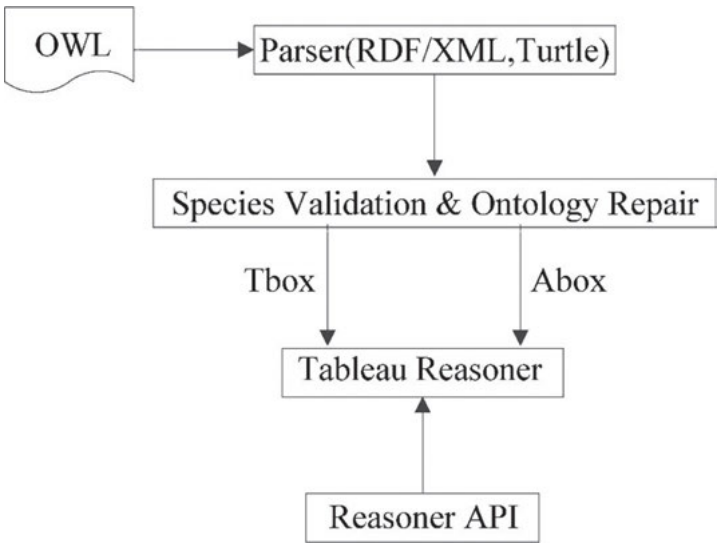

Figure 7. The workflow of the Pellet inference engine.

2) Firm(?f)^ROE(?r)^hasFinancialVariable(?f,?r)^has Value(?r,?v1)^swrlb:greaterTh an(?v1,-0.0433)^Reve nue(?e)^hasFinancialVariable(?f,?e)^hasValue(?e,?v2) ^swr lb:lessThanOrEqual(?v2,12650.0)^ProfitMar gin(?p)^hasFinancialVariable(?f,?p)^has Value(?p,?v3)^swrlb:lessThanOrEqual(?v3,0.3625) ^CashFlow/OperatingRevenue(?o)^hasFinancial Variable(?f,?o)^hasValue(?o,?v4)^swrlb:lessThanOr Equal(?v4,0.3505)->hasFraudActivity(?f,true);
3) Firm(?f)^ROE(?r)^hasFinancialVariable(?f,?r)^has Value(?r,?v1)^swrlb:greaterThan(?v1,-0.0433)^Revenue (?e)^hasFinancialVariable(?f,?e)^hasValue(?e,?v2)^swr lb:greaterThan(?v2,12650)^swrlb:lessThanOrEqual (?v2,2630000)^ProfitMargin(?p)^hasFinancialVariable (?f,?p)^hasValue(?p,?v3)^swrlb:lessThanOrEqual (?v3,0.3625)^CashFlow/OperatingRevenue(?o)^has FinancialVariable(?f,?o)^hasValue(?o,?v4)^swrlb:less ThanOrEqual(?v4,0.3505)^NetAssetsTurnover(?n) ^hasFinancialVariable(?f,?n)^hasValue(?n,?v5)^swrlb: lessThanOrEqual(?v5,1.65)->hasFraudActivity (?f,false);
4) Firm(?f)^ROE(?r)^hasFinancialVariable(?f,?r)^has Value(?r,?v1)^swrlb:greaterThan(?v1,-0.0433)^Revenue (?e)^hasFinancialVariable(?f,?e)^hasValue(?e,?v2)^swr lb:greaterThan(?v2,12650)^swrlb:lessThanOrEqual (?v2,2630000)^ProfitMargin(?p)^hasFinancialVariable (?f,?p)^hasValue(?p,?v3)^swrlb:lessThanOrEqual (?v3,0.3625)^CashFlow/OperatingRevenue(?o)^has FinancialVariable(?f,?o)^hasValue(?o,?v4)^swrlb:less ThanOrEqual(?v4,0.3505)^NetAssetsTurnover(?n)^ hasFinancialVariable(?f,?n)^hasValue(?n,?v5)^swrlb: greaterThan(?v5,1.65)->hasFraudActivity(?f,true);
5) Firm(?f)^ROE(?r)^hasFinancialVariable(?f,?r)^has Value(?r,?v1)^swrlb:greaterThan(?v1,-0.0433)^Revenue (?e)^hasFinancialVariable(?f,?e)^hasValue(?e,?v2)^swr lb:lessThanOrEqual(?v2,2630000)^ProfitMargin(?p)^ hasFinancialVariable(?f,?p)^hasValue(?p,?v3)^swrlb: lessThanOrEqual(?v3,0.3625)^CashFlow/Operating Revenue(?o)^hasFinancialVariable(?f,?o)^hasValue (?o,?v4)^swrlb:greaterThan(?v4,0.3505)->hasFraud Activity(?f,false);
6) Firm(?f)^ROE(?r)^hasFinancialVariable(?f,?r)^has Value(?r,?v1)^swrlb:greaterThan(?v1,-0.0433)^Revenue (?e)^hasFinancialVariable(?f,?e)^hasValue(?e,?v2)^swr lb:lessThanOrEqual(?v2,2630000)^ProfitMargin(?p) ^hasFinancialVariable(?f,?p)^hasValue(?p,?v3)^swrlb: greaterThan(?v3,0.3625)->hasFraudActivity(?f,true);
7) Firm(?f)^ROE(?r)^hasFinancialVariable(?f,?r)^has Value(?r,?v1)^swrlb:greaterThan(?v1,-0.0433)^Revenue (?e)^hasFinancialVariable(?f,?e)^hasValue(?e,?v2)^swr lb:greaterThan(?v2,2630000)^swrlb:lessThanOrEqual (?v2,7465000)^NetAssetsTurnover(?n)^hasFinancial Variable(?f,?n)^hasValue(?n,?v3)^swrlb:lessThanOr Equal(?v3,3.695)->hasFraudActivity(?f,false);
8) Firm(?f)^ROE(?r)^hasFinancialVariable(?f,?r)^has Value(?r,?v1)^swrlb:greaterThan(?v1,-0.0433)^Revenue (?e)^hasFinancialVariable(?f,?e)^hasValue(?e,?v2)^swr lb:greaterThan(?v2,2630000)^swrlb:lessThanOrEqual (?v2,7465000)^NetAssetsTurnover(?n)^hasFinancial Variable(?f,?n)^hasValue(?n,?v3)^swrlb:greaterThan (?v3,3.695)->hasFraudActivity(?f,true);
9) Firm(?f)^ROE(?r)^hasFinancialVariable(?f,?r)^has Value(?r,?v1)^swrlb:greaterThan(?v1,-0.0433)^Revenue (?e)^hasFinancialVariable(?f,?e)^hasValue(?e,?v2)^swr lb:greaterThan(?v2,7465000)->hasFraudActivity (?f,true).

5.0 Experiment and discussion

5.1 Construction of a financial statement fraud detection system

In this study, we used Protégé 5.2 to build a financial statement fraud detection ontology. Figure 4 shows the Protégé ontology editor. The "owl: Thing" has three subclasses: financial statement, financial variables, and firm. In the financial variables class, five subclasses are defined, and each subclass has sixty instances. In the firm class, thirty-one subclasses are defined and sixty firm instances are contained in those subclasses based on the SIC code. The number of instances in the financial statement fraud detection ontology is 360.

SWRL Tab is a Protégé plug-in where SWRL rules can be edited. Figure 8 shows the rules that are edited in SWRL Tab. Through the implementation of the Pellet inference engine, it can be seen that sixty axioms have been added. As shown in Figure 9, the instance "ComcastCorp2000" has its attribute "hasFraudActivity" assigned the value "true." This shows the validity of the inference.

In this experiment, 200 firms' financial statements were used in C4.5 decision tree generation and fraud detection rules extraction. Sixty firms and firms' financial statements were used for fraud detection in a knowledge-based financial statement fraud detection system. Twenty-six firms' datatype properties "hasFraudActivity" were assigned the value "false." Thirty-four firm's datatype properties "hasFraudActivity" were assigned the value "true."

5.2 Performance metrics

Evaluation of the performance of fraud detection system is an important step. We used accuracy, TP rate (also called recall rate), and F-measure to evaluate the performance of the system in this paper. In Table 3, some parameters are defined and explained, again using "fraudulent" and "non-fraudulent" as shorthand for firms that commit or do not commit fraud.

In this study, TP equals twenty-six, namely, the number of fraudulent firms detected as fraudulent is twenty-six. TN equals twenty-two, namely, the number of non-fraudulent firms detected as non-fraudulent is twenty-two. FP equals eight, namely, the number of non-fraudulent firms incorrectly detected as fraudulent is eight. FN equals four, namely, the number of fraudulent firms incorrectly detected as non-fraudulent is four.

Table 4 shows the detection system's performance results. In sixty firms, the system achieved 80% accuracy, a recall rate of 86.67% and a F-measure of 78.2%. In this detection system, all detection rules are extracted from machine learning algorithms, and no domain expert was involved in the experiment. This result shows the validity of the system.

Accuracy = (TP+TN)/(P+N)	TP is the number of fraudulent firms detected as fraudulent. TN is the number of non-fraudulent firms detected as non-fraudulent. P is the number of fraudulent firms and N is the number of non-fraudulent firms.
TP rate = TP/P	TP rate (also called recall rate) is the percentage that number of all fraudulent firms divided by number of firms correctly detected as fraudulent.
TN rate = TN/N	TN rate is the percentage that number of all non-fraudulent firms divided by number of firms correctly detected as non-fraudulent.
FP rate= FP/N	FP rate is the percentage that number of all non-fraudulent firms divided by number of firms incorrectly detected as fraudulent.
FN rate = FN/P	FN rate is the percentage that number of all fraudulent firms divided by number of firms incorrectly detected as non-fraudulent.
F-measure = (2*Precision*TP rate)/(Precision + TP rate)	F-measure is the harmonic mean of precision and TP rate.

Table 3. Performance evaluation of financial statement fraud detection system.

Accuracy	TP rate	F-measure
80.00%	86.67%	78.20%

Table 4. Evaluation results of financial statement fraud detection system.

6.0 Conclusion and future work

Financial fraud is an important issue that widely concerns the financial industry and academia. Financial fraud can reduce the trust of stock market participants in the market and cause serious economic problems. Financial statement fraud, a typical fraud activity in financial fraud, has caused several bankruptcies and huge economic loss in the last two decades. Thus, the detection of financial statement fraud, the discovery of fraud patterns, and the improvement of fraud detection efficiency have become important topics in the industry and in academia.

Our study presents a knowledge-based financial statement detection system by using a machine-learning algorithm to discover the financial variables and fraud detection rules and using an ontology and inference engine to discover implicit knowledge. To select informative features, we per-

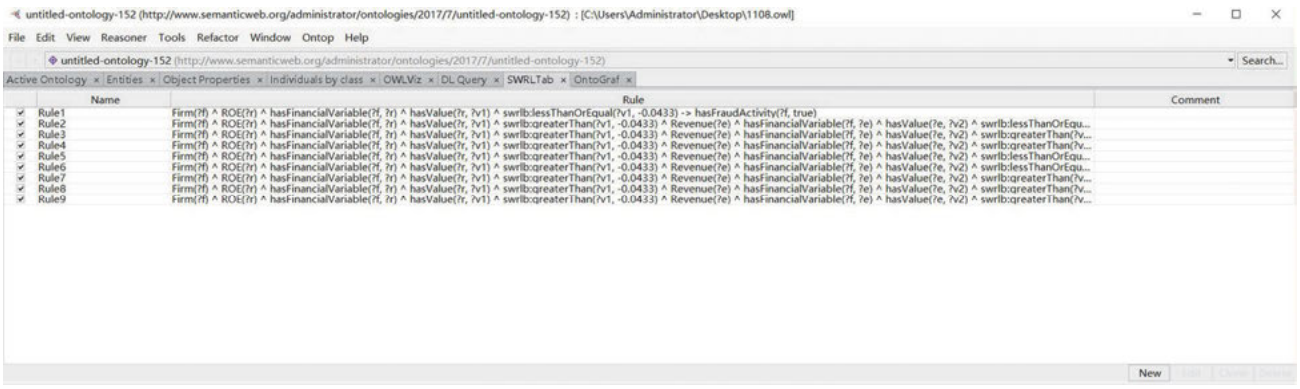

Figure 8. SWRL rule editing.

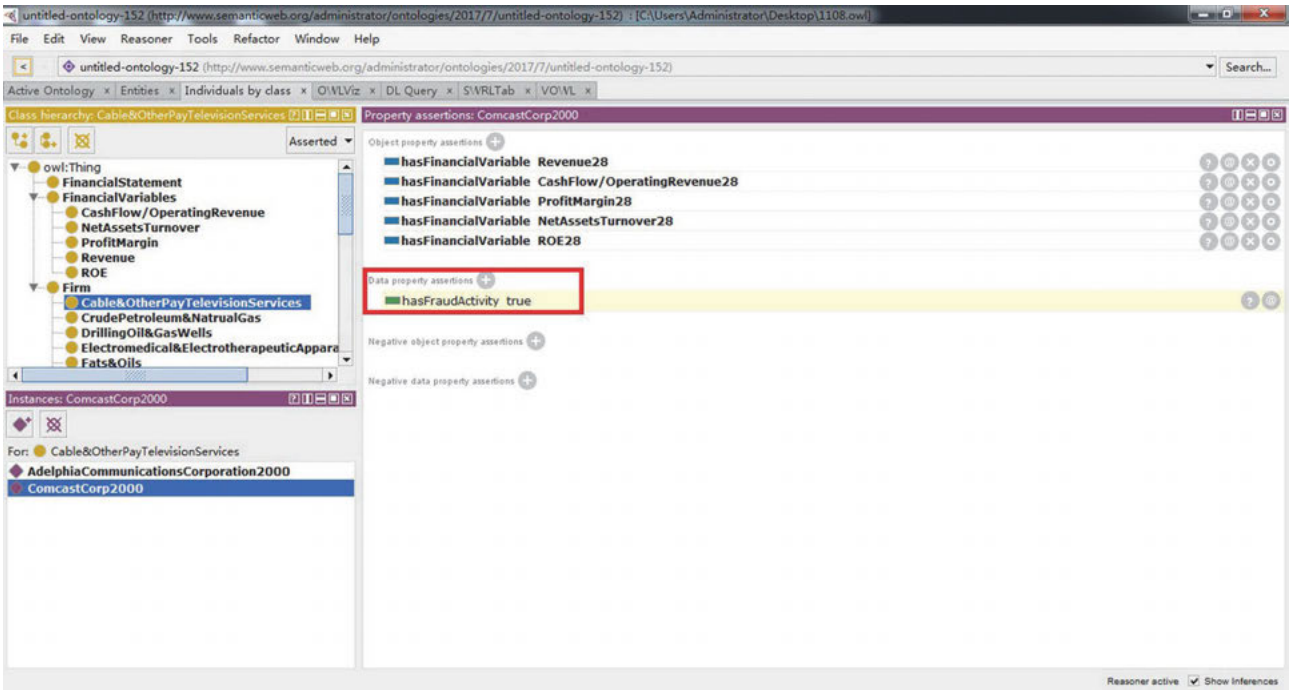

Figure 9. Inference result of financial statement fraud detection ontology.

formed feature selection by using an extremely randomized trees algorithm. In this study, we use OWL to construct a financial statement fraud detection ontology. We employed a C4.5 decision tree to extract financial statement fraud detection rules and used SWRL and OWL to describe the semantic information of decision rules. OWL and SWRL explain the relationships among financial statements, financial variables, and fraud activity at a semantic level. The inference engine was employed to utilize existing knowledge and discover new knowledge. This study identifies financial statement fraud at the semantic level and provides a method for semi-automatic ontology construction. It illustrates another method for the construction of similar ontologies in other domains. Future studies could improve the ontological knowledge of financial statement detection and discover more rules by using machine-learning algorithms and expert knowledge. Besides financial variables, future work could also focus on the textual content of financial statements to find more semantic information to improve the efficiency of financial statement fraud detection.

References

Abbasi, Ahmed, Conan Albrecht, Anthony Vance, and James Hansen. 2012. "Metafraud: A Meta-Learning Framework for Detecting Financial Fraud." *MIS Quarterly* 36, no.4: 1293-1327.

Abburu, Sunitha. 2012. "A Survey on Ontology Reasoners and Comparison." *International Journal of Computer Applications* 57, no. 17: 33-9.

Beasley, Mark S., Joseph V. Carcello, Dana R. Hermanson, and Terry L. Neal. 2010. "Fraudulent Financial Report-

ing: 2998-3007; An Analysis of U. S. Public Companies Research." New York: Committee of Sponsoring Organizations of the Treadway Commission (COSO). https://www.coso.org/Documents/COSO-Fraud-Study-2010-001.pdf

Bennett, Mike. 2013. "The Financial Industry Business Ontology: Best Practice for Big Data." *Journal of Banking Regulation* 14: 255-68.

Bell, T. B., S. Szykowny, and J. J. Willingham. 1991. "Assessing the Likelihood of Fraudulent Financial Reporting: A Cascaded Logit Approach." *Journal of Accounting and Economics* 26: 475-500.

Campos, Maria Luiza de Almeida and Hagar Espanha Gomes. 2017. "Ontology: Several Theories on the Representation of Knowledge Domains." *Knowledge Organization* 44: 178-86.

Cecchini, Mark, Haldun Aytug, Gary J. Koehler, and Praveen Pathak. 2010. "Detecting Management Fraud in Public Companies." *Management Science* 56: 1146-60.

Chen, Suduan, Yeong-Jia James Goo, and Zone-De Shen. 2014 "A Hybrid Approach of Stepwise Regression, Logistic Regression, Support Vector Machine, and Decision Tree for Forecasting Fraudulent Financial Statements." *Scientific World Journal.* doi:10.1155/2014/968712

Dechow, Patricia M., Weili Ge, Chad R. Larson, and Richard G. Sloan. 2011. "Predicting Material Accounting Misstatements." *Contemporary Accounting Research* 28: 17-82.

Dong, Wei, Stephen Shaoyi Liao, Bing Fang, Xian Cheng, Zhu Chen, and Wenjie Fan. 2014. "The Detection of Fraudulent Financial Statements: An Integrated Language Model." *Proceedings of 19th Pacific Asia Conference on Information Systems, June 24-28, 2014, Chengdu, China.* Atlanta, GA: Association for Information Systems AIS eLibrary, 383. https://aisel.aisnet.org/pacis2014/383

Dyck, Alexander, Adair Morse, and Luigi Zingales. 2010. "Who Blows the Whistle on Corporate Fraud?" *Journal of Finance* 65: 2213-53.

Geurts, Pierre, Damien Ernst, and Louis Wehenkel. 2006. "Extremely Randomized Trees." *Machine Learning* 63: 3-42.

Gruber, Thomas R. 1993. "A Translation Approach to Portable Ontology Specifications." *Knowledge Acquisition* 5: 199-220.

Guyon, Isabelle and André Elisseeff. 2003. "An Introduction to Variable and Feature Selection." *Journal of Machine Learning Research* 3: 1157-82.

Han, Jiawei, Jian Pei, and Micheline Kamber. 2011. *Data Mining: Concepts and Techniques*, 3rd ed. Burlington, MA: Elsevier.

Hajek, Petr and Roberto Henriques. 2017. "Mining Corporate Annual Reports for Intelligent Detection of Financial Statement Fraud: A Comparative Study of Machine Learning Methods." *Knowledge-Based Systems* 128: 139-52.

Hastie, Trevor, Robert Tibshirani, and Jerome Fridman. 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Series in Statistics. New York: Springer.

Horrocks, Ian, Peter F. Patel-Schneider, and Frank Van Harmelen. 2003. "From SHIQ and RDF to OWL: The Making of a Web Ontology Language." *Web Semantics: Science, Services and Agents on the World Wide Web* 1: 7-26. doi:10.1016/j.websem.2003.07.001

Horrocks, Ian, Peter F. Patel-Schneider, Harold Boley, Said Tabet, Benjamin Grosof, and Mike Dean. 2004. "SWRL: A Semantic Web Rule Language Combining OWL and RuleML." https://www.w3.org/Submission/2004/SUBM-SWRL-20040521/

Kanellopoulos, Dimitris, Sotiris Kotsiantis, and Vasilis Tampakas. 2007. "Towards an Ontology-Based System for Intelligent Prediction of Firms with Fraudulent Financial Statements." In *ETFA 2007: 12th IEEE International Conference on Emerging Technologies and Factory Automation: ETFA 2007 Proceedings; September 25-28, 2007 University of Patras, Greece.* [Piscataway, NJ]: Institute of Electrical and Electronics Engineers, 1300-7. doi:10.1109/EFTA.2007.4416931

Kingston, John, Burkhard Schafer, and Wim Vandenberghe. 2004. "Towards a Financial Fraud Ontology: A Legal Modelling Approach." *Artificial Intelligence and Law* 12, no. 4: 419-46.

Kirkos, Efstathios, Charalambos Spathis, and Yannis Manolopoulos. 2007. "Data Mining Techniques for the Detection of Fraudulent Financial Statements." *Expert Systems with Applications* 32: 995-1003.

Kohavi, Ron. 1995. "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." In *IJCAI-95: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, Montréal, Québec, Canada, August 20-25, 1995*, ed. Chris S. Mellish. San Mateo, CA: Morgan Kaufmann, 1137-45.

Kotsiantis, S., E. Koumanakos, D. Tzelepis, and V. Tampakas. 2006. "Forecasting Fraudulent Financial Statements Using Data Mining." *International Journal of Computational Intelligence* 3, no. 2: 104-10.

Li, Xinyang, Wei Xu and Xuesong Tian. 2014. "How to Protect Investors? A GA-based DWD Approach for Financial Statement Fraud Detection." In *Proceedings 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC) October 5-8, 2014 San Diego, CA, USA.* [Piscataway, NJ]: Institute of Electrical and Electronics Engineers, 3548-54. doi:10.1109/SMC.2014.6974480

Liu, Chengwei, Yixiang Chan, Syed Hasnain Alam Kazmi, and Hao Fu. 2015. "Financial Fraud Detection Model:

Based on Random Forest." *International Journal of Economics and Finance* 7, no. 7: 178-88.

Liu, Tianyu. 2007 "The Research of Ensemble Learning and its Application Based on Feature Selection." PhD diss., Shanghai University.

McGuinness, Deborah L. and Frank van Harmelen. 2004. "OWL Web Ontology Language Overview." https://www.w3.org/TR/2004/REC-owl-features-20040210/

Noy, Natalya Fridman, Ray W. Fergerson, and Mark A. Musen. 2000. "The Knowledge Model of Protege-2000: Combining Interoperability and Flexibility." In *Knowledge Engineering and Knowledge Management Methods, Models, and Tools; 12th International Conference, EKAW 2000 Juan-les-Pins, France, October 2-6, 2000 Proceedings*, ed. Rose Dieng and Olivier Corby. Lecture Notes in Computer Science 1937. Berlin: Springer, 17-32.

O'Connor, Martin, Holger Knublauch, Samson Tu, Benjamin Grosof, Mike Dean, William Grosso, and Mark Musen. 2005. "Supporting Rule System Interoperability on the Semantic Web with SWRL." In *The Semantic Web—ISWC 2005: 4th International Semantic Web Conference, ISWC 2005, Galway, Ireland, November 6-10, 2005; Proceedings* ed. Yolanda Gil, Enrico Motta, V. Richard Benjamins, and Mark A. Musen. Lecture Notes in Computer Science 3729. Berlin: Springer, 974-86. doi:10.1007/11574620_6

Omar, Normah, Zulaikha 'Amirah Johari, and Malcolm Smith. 2017. "Predicting Fraudulent Financial Reporting Using Artificial Neural Network." *Journal of Financial Crime* 24: 362-87.

Opitz, David W. 1999. "Feature Selection for Ensembles." In *Proceedings: Sixteenth National Conference on Artificial Intelligence (AAAI-99); Eleventh Innovative Applications of Artificial Intelligence Conference (IAAI-99).* Menlo Park, CA: AAAI Press, 379-84. https://www.aaai.org/Papers/AAAI/1999/AAAI99-055.pdf

Padmavathi, Thimmaiah and Madaiah Krishnamurthy. 2017. "Semantic Web Tools and Techniques for Knowledge Organization: An Overview." *Knowledge Organization* 44: 273-90.

Pai, Ping-Feng, Ming-Fu Hsu, and Ming-Chieh Wang. 2011. "A Support Vector Machine-Based Model for Detecting Top Management Fraud." *Knowledge-Based Systems* 24: 314-21.

Pan, Jeff Z., Giorgos Stoilos, Giorgos Stamou, Vassilis Tzouvaras, and Ian Horrocks. 2005. "f-SWRL: A Fuzzy Extension of SWRL." In *Formal Models and Their Applications.* Part II of *Artificial Neural Networks: ICANN 2005;15th International Conference, Warsaw, Poland, September 11-15, 2005; Proceedings*, ed. Wlodzislaw Duch, Erkki Oja, and Slawomir Zadrozny. Lecture Notes in Computer Science 3697. Berlin: Springer, 829-34. doi: 10.1007/11550907_131

Ravisankar, P., V. Ravi, G. Raghava Rao, and I. Bose. 2011. "Detection of Financial Statement Fraud and Feature Selection Using Data Mining Techniques." *Decision Support Systems* 50: 491-500.

Quinlan, J. R. 1987. "Generating Production Rules from Decision Trees." In *IJCAI 87: Proceedings of the Tenth International Joint Conference on Artificial Intelligence, August 23-28, 1987*. Los Altos, CA: Morgan Kaufmann, 304-7

Quinlan, J. R. 1993. *C4.5: Programs for Machine Learning*. The Morgan Kaufmann Series in Machine Learning. San Mateo, CA: Morgan Kaufmann.

Rezaee, Zabihollah. 2005. "Causes, Consequences, and Deterrence of Financial Statement Fraud." *Critical Perspectives on Accounting* 16: 277-98.

Shue, Li-Yen, Ching-Wen Chen, and Weissor Shiue. 2009. "The Development of an Ontology-Based Expert System for Corporate Financial Rating." *Expert Systems with Applications.* 36: 2130-42.

Sirin, Evren, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur, and Yarden Katz. 2007. "Pellet: A Practical Owl-Dl Reasoner." *Web Semantics: Science, Services and Agents on The World Wide Web* 5: 51-3.

Song, Xin-Ping, Zhi-Hua Hu, Jian-Guo Du, and Zhao-Han Sheng. 2014. "Application of Machine Learning Methods to Risk Assessment of Financial Statement Fraud: Evidence from China." *Journal of Forecasting* 33: 611-26. doi:10.1002/for.2294

Summers, Scott L. and John T. Sweeney. 1998. "Fraudulently Misstated Financial Statements and Insider Trading: An Empirical Analysis." *Accounting Review* 73: 131-46.

West, Jarrod and Maumita Bhattacharya. 2016. "Intelligent Financial Fraud Detection: A Comprehensive Review." *Computers & Security* 57: 47-66.

Zhao, Gang, John Kingston, Koen Kerremans, Frederik Coppens, Ruben Verlinden, Rita Temmerman, and Robert Meersman. 2004. "Engineering an Ontology of Financial Securities Fraud." In *On the Move to Meaningful Internet Systems 2004: OTM 2004 Workshops; OTM Confederated International Workshops and Posters, GADA, JTRES, MIOS, WORM, WOSE, PhDS, and INTEROP 2004, Agia Napa, Cyprus, October 25-29, 200; Proceedings,* ed. Robert Meersman, Zahir Tari, and Angelo Corsaro. Lecture Notes in Computer Science 3292. Berlin: Springer, 605-20. doi: 10.1007/978-3-540-30470-8_73