# FINANCIAL FRAUD RECOGNITION AND IDENTIFICATION METHOD USING REASONING AND QUANTITATIVE ASSOCIATION EVALUATION

**Mariusz Chmielewski, Ph.D.**
*Cybernetics Department*
*Military University of Technology*
Warsaw, Poland
*e-mail: mchmielewski@wat.edu.pl*

**Maciej Kiedrowicz, Ph.D.**
*Cybernetics Department*
*Military University of Technology*
Warsaw, Poland
*e-mail: maciej.kiedrowicz@wat.edu.pl*

**Piotr Stąpor, M.Sc.**
*Cybernetics Department*
*Military University of Technology*
Warsaw, Poland
*e-mail: piotr.stapor@wat.edu.pl*

**Abstract**

This work presents, a tool and a method dedicated for financial fraud recognition and evaluation based on stream of transactional data from financial institutions. Presented novel approach aims at context-aware data processing coming from ontology application and reasoning capabilities utilising DL and FOL (BAADER et al. 2003; STAAB et al., 2004) reasoners. Domain terminology defines elemental concepts, relations and classification rules which provide semantic processing capabilities. This article summarizes the proof of implementation of the concept of the IAFEC Ontology Toolkit for identifying fraud based on a set of problem solving ontologies. Methods, algorithms and software are inputs to IAFEC analytical tools demonstrating semantic-aware association analysis. A novelty in this approach is the inclusion of heterogeneous data analysis, which combine various layers of data extending range of available associations between individuals, organisations and financial market actors. Rich domain descriptions provide multiple ways of expressing relations in families, social groups, organisations, financial transactions and other dependencies. Progress in automatic reasoning and the availability of semantic processing tools (e.g. Protégé, Jena) encourage analyst to extend existing link analysis methods to contextual knowledge processing. Presented research provides, a high level insight into the analytical method and algorithms, which are based on logical reasoning, identification and ranking of associations found in financial transactions supplemented with intelligence data. Elaborated method is delivered as standalone desktop application integrated with Protégé OWL 5.0 (PROTÉGÉ WEB, 2016) reasoners and data integration services. Tool's architecture simplifies integration with available semantic processing plug-ins while delivering functionalities for analytical process definition.

**Key words:** financial fraud identification, money laundering analytics, knowledge discovery, semantic association, ontologies, context-aware processing

## Introduction

Money laundering is becoming a key concern for governments because of its obvious impact on tax evasion and the development of organised crime. The involvement of technology and the globalisation of remittances make it even more difficult to identify and recognise such practices. For many years, EU governments have been using ICT tools to monitor financial transfers to detect suspicious operations. However, this analysis is mainly based on manually implemented rules and expertise. The same is true of new cases of fraud or deception camouflaged by organisational or individual chains, facts relating to production time or loosely. In such cases, there is a need for tools for filtering and ranking automatic data, which can operate at the initial stage of processing and retrieving transfers from the financial data stream. The complexity of the problems also generates a need for research and development in the field of automatic extraction, classification and reasoning of data. Thorough analysis of **s**everal research projects

demonstrated various approaches to construct and model financial domain and provide mechanisms for information processing and deducting fraud, such as FF POIROT (BASILI et al., 2003), DOGMA (JARRAR, 2008). As part of this work, the available research materials were reviewed and valuable semantic modelling assumptions have been adopted (CHMIELEWSKI, 2009; BASILI et al., 2003; JARRAR, 2008).

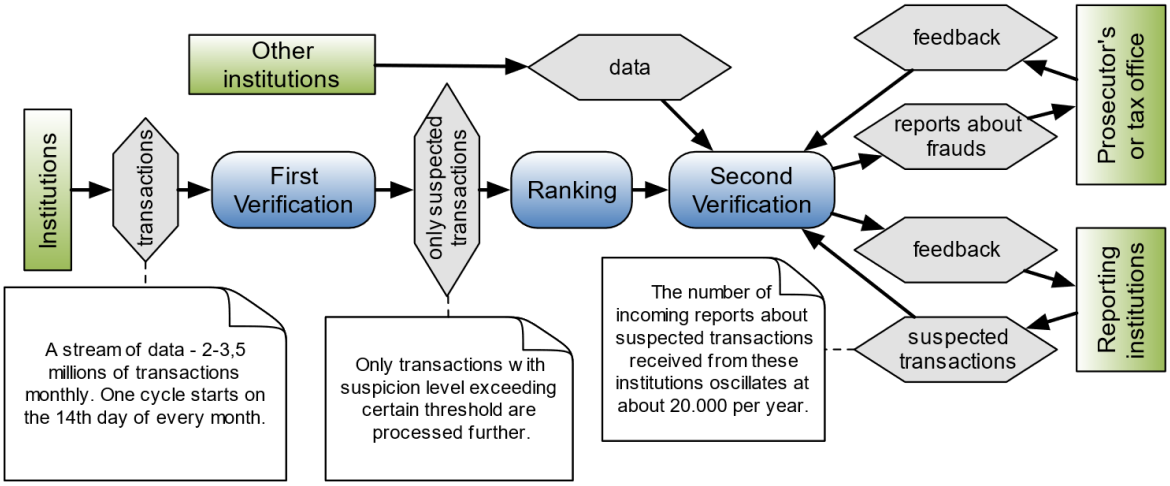


**Fig. 1.** Integrated financial data flow and fraud detection process
(processed by financial transaction monitoring organisations - GIIF).
*Source: Own study.*

The IAFEC project combines several analytical methods for fraud detection, mainly related to the use of network analysis, reasoning and structural analysis (provided in form of quantitative nod qualitative methods). A novelty of the presented approach is the use of domain modelling and probing ontology in the data processing chain in order to identify hidden associations (hidden relationships) in the knowledge base processed by DL, FOL (BAADER et al., 2003) and semantic graph processing reasoners. The use of semantic models also gives another advantage, the implementation of contextual-conscious relations, extending the possibilities of multi-perspective methods.

**Regulations and method requirements**

The General Inspector of Financial Information (GIIF) in Poland is responsible for preventing money laundering and terrorist financing. (PMoF,2012). GIIF collects large number of transaction records, ranging from 2 to 3.5 million per month. The data undergo a first stage verification, where it is classified as normal or suspect. Transactions marked as belonging to the latter category are then classified according to the level of presumption of fraud since the first verification, the amount of money involved and the importance (standard fraud, terrorism, corruption in government, etc.) of the transaction. Transactions with a score above the threshold shall be further processed.

The following sources of information were analysed in the task (PMoF, 2012):

- (REGULATION OF THE FINANCIAL MINISTER) regarding the transaction register and the mode of providing information to the General Inspector for Financial Information,
- Operating instructions for the transaction log functions for the General Inspector for Financial Information in the FAKT program,
- Operating Instructions for the Transaction Register Program (GIIF),
- Scientific material relating to financial transactions ,
- ontologies of financial crime (UCOM, FF POIROT, DOGMA),
- ontologies of family and business relationships (folksonomies).

There are many criteria for suspicious operations: sometimes it is the value of the operations, their frequency or the unusual place from which they are carried out (a history of financial operations and an analysis of the anomalies of deviations from the normal rule are required). Generally, it is about financial operations carried out through various channels, which differ from the previous behaviour of the client (behavioural analysis). For security reasons, such transactions may be additionally confirmed by telephone. Therefore, it is very important to update contact details and, above all, telephone numbers in the banking system on an ongoing basis.

According to the legal acts, since 01.01.2004 notaries and financial institutions: banks or broker houses, have to register for inspection all transactions exceeding 15 000 Euro and forward them to the GIIF. Regardless of the value of the operation, so-called suspicious transactions, i.e. transactions whose circumstances indicate that we are dealing with money laundering, must also be recorded. Related transactions that consist of several operations are also recorded where circumstances indicate that they are linked and where individual payments are usually below a threshold amount. Recorded transactions (suspicious transactions) consist of participating entities data, amount and time as well as a type of operation (channel). At the written request of the General Inspector for Financial Information (GIIF), the heads of an obliged institution, which received an order from its client to carry out a specific transaction (which is suspected to have been a crime under Article 299 of the Penal Code), may withhold such a transaction or block the account. A decision to block an account for which financial operations have been executed must be issued by a public prosecutor – the penalties of such crimes include imprisonment up to 8 years. Entrepreneurs who run companies classified as so called obliged must take into account the fact that GIIF employees may conduct audit without preliminary schedule. The GIIF auditors have the right of access to the premises of the audited institution, and view the financial documents, their copies and other materials, which may serve as evidence of the case, in an unlimited way. They may request oral and written explanations from the owner and employees of the institution concerning their activities.

The second stage of verification process involves access to additional data (e.g. registers from the Legal Register of Enterprises in Poland - KRS, data compiled by the Central Statistical Office in Poland, information from CEPIK (Central Register of Vehicles and Drivers) or CEIDG (Central Register and Information on Business Activity)) and information provided by external institutions, which are obliged to provide them upon request of the GIIF, may be taken into account. The two main problems at this stage are incompatible data formats received from third parties and the inability to process all suspicious transactions within a reasonable time. In Poland, there are a number of financial institutions which may submit suspicious transactions directly to the General Inspectorate for Financial Information (GIIF). These records bypass the first part of the process and are automatically accepted for the second stage verification. When fraud is detected, the case is reported to the prosecution or tax office and feedback is expected from the latter to help improve the verification process. The GIIF provides also feedback to the reporting institutions in order to improve their pre-filtration capacity.

**Financial data analysis method implementing semantic representations**

The presented approach addresses the problem of storing and analysing a large volume of asynchronous financial transactions. For reader's convenience a schema of the entire process has been attached below.

The system loads the definitions of terminology from a dedicated ontology. After that the system is ready to read data and map the transactions' records to instances placed in ontology. These individuals are then assigned to respective concepts and connected with necessary relations between themselves and instances already contained within knowledgebase. The aforementioned preprocessing is done via web-services, that contain transformation mechanisms implemented in Java. Based on conducted tests no available reasoner is able to efficiently process that amount of data in transactional mode. In response to this problem a following solution was devised. Transactions are aggregated into sets during preprocessing with respect to time of realization and transaction sides (Each set is represented as an individual with datatype roles attached, pointing to numerical indicators – e.g. total amount value). The instance base is transformed into multigraph and divided into connected components, of which each one undergoes reasoning separately. (Lack of path between vertices indicates a lack of semantic linkage between the individuals they represent. A possibility, albeit small, that some connection between such pair could still exist is conceivable. Given circumstances – problem size and complexity, however, authors believe such approach is justified and permissible).

SWRL rules on such a large dataset is costly in terms of time and memory. This has been solved by running a set of aggregation procedures which increases further data processing and inference efficiency. These algorithms have been implemented in Java programming language. The rules they embody, search for certain structure patterns and generate new individuals, each one representing a set of aggregated data and each one equipped with new datatype properties links. These arcs point to values obtained by applying various math functions for individuals' datatype values in the analysed set. E.g. the algorithm sums all transactions done by a given company (represented as datatype relations) in a month and produces as a result an object of OWL class HighValueTransactionSet with values such as total and average transaction values. The aggregated data fuels reasoning engine of main fraud detecting ontology, during which HermiT classifies individuals into concepts by executing its hyper-tableau algorithm and connects (via object-type properties) them according to our set of DL-safe SWRL-defined rules (BAADER et al. 2003). The end of

HermiT's work marks the start of semantic structure analysis. A devised, dedicated tool transforms instances' base into a multigraph and then divides it into a set of weak connected components. of which each one undergoes structural reasoning separately. The tool evaluates various coefficients for vertices, edges and directly not connected pairs of individuals (given, that they belong to the same weak connected component). This approach enables to greatly reduce this stage's running time. Of course, a pair of vertices from different components can in reality be associated, but a probability that such semantic linkage exists is considerably negligible given problem's size and complexity (BARTHELEMY et al., 2005).

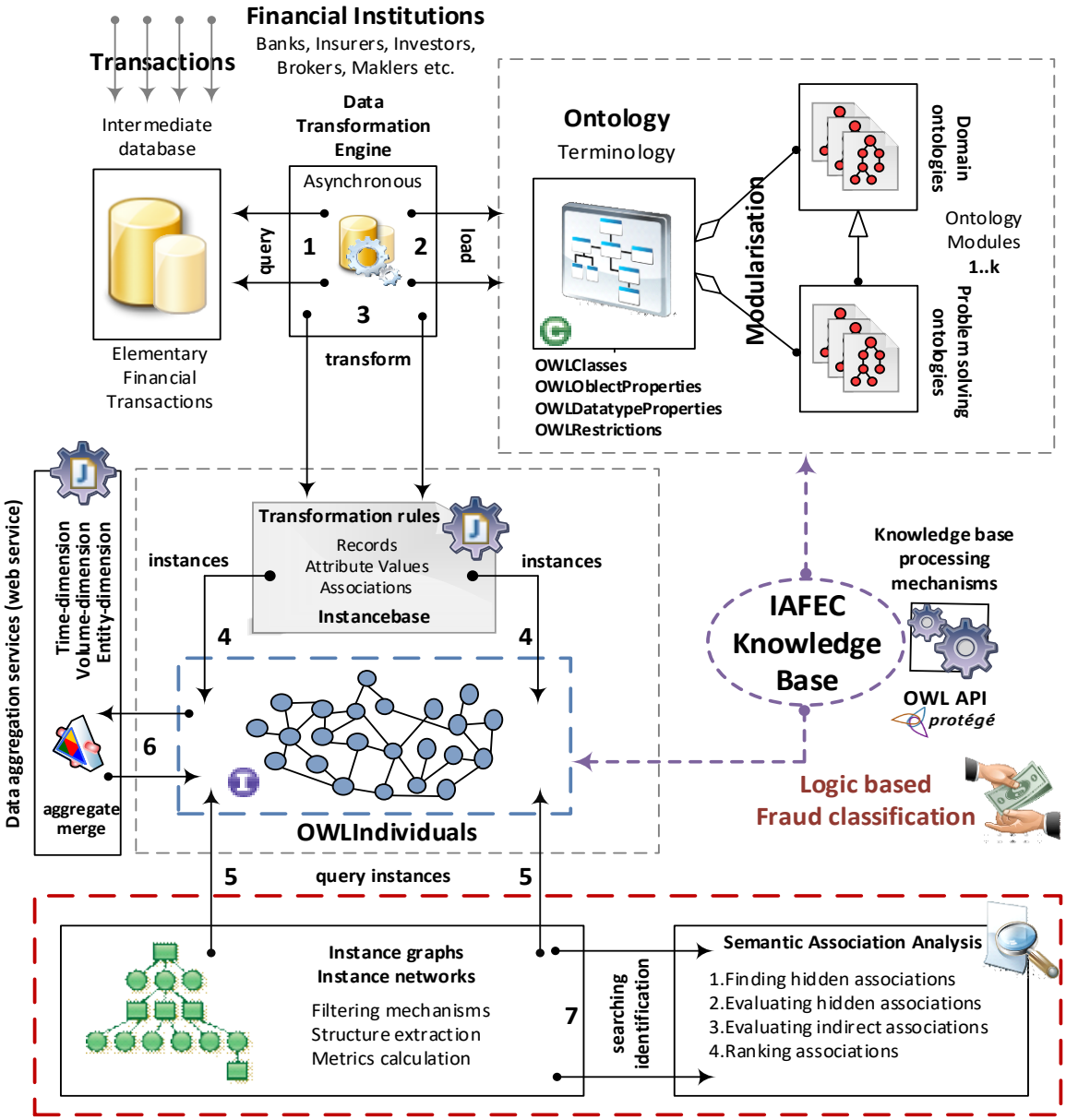


**Fig. 2.** The proposed method scheme. Processing stages start from elementary data, semantic data enrichment, aggregation, quantitative evaluation to reasoning. The heterogeneous data and context-aware relationships processing is a result of 1-7.
*Source: Own study.*

Because of resulting multigraphs' magnitude, it might be both difficult and time consuming for user to consume the entire output. It is natural to expect that within such large structure the importance levels of some discoveries prevail over others. The classification of certain company's instance to a *HighlySuspiciousCompany* concept is undoubtedly more significant than classification of *TransactionSide* as *FinancialInstitution*, but the problem that remains is: how to bring to an attention of a user some meaningful connections realized as chains of instances joined by object-properties links. One of possible ways to deal

with it is by the application of semantic association ranking method, described in (CHMIELEWSKI, 2009). Review chapter 5.2 for details on structural analysis and ontology transformation rules.

**Semantic descriptions in IAFEC Ontology utilised in association detection**

A number of ontological models, has been developed covering terminology directly or remotely connected to financial operations and market. IAFEC Foundations is a set of business concepts that are designed to support the semantics of the financial industry's terms and conditions as set out in other specifications. The IAFEC Foundations models define concepts that are not unique to the financial services industry. On this basis, financial industry terms in other IAFEC specifications can be obtained by extensions. It shall also include terms to which the characteristics of the items may refer in those specifications. IAFEC foundations therefore contain a number of basic concepts concerning, among others, legal, contractual and organisational concepts. The content of the ontology (GRUBER, 1993; STAAB et al., 2004) is a documentation, interpretable in formal terms, of the concepts represented by the financial industry concepts used in official documents of financial organizations, such as contracts, product/service specifications, and documents relating to compliance with laws and regulations. Another part of modelled terminology concerns tax frauds (VAT within the EU) and investment fraud (online) which cover:

- stock market (operations and exceptions) counterfeiting in relation to securities;
- impersonating natural or legal persons allegedly having affiliations with genuinely existing legal financial institutions;
- illegal insider trading (by existing relations with suspected organisations) - fraud and manipulation involving the sale or sharing of company information between employees (insiders) and outsiders.

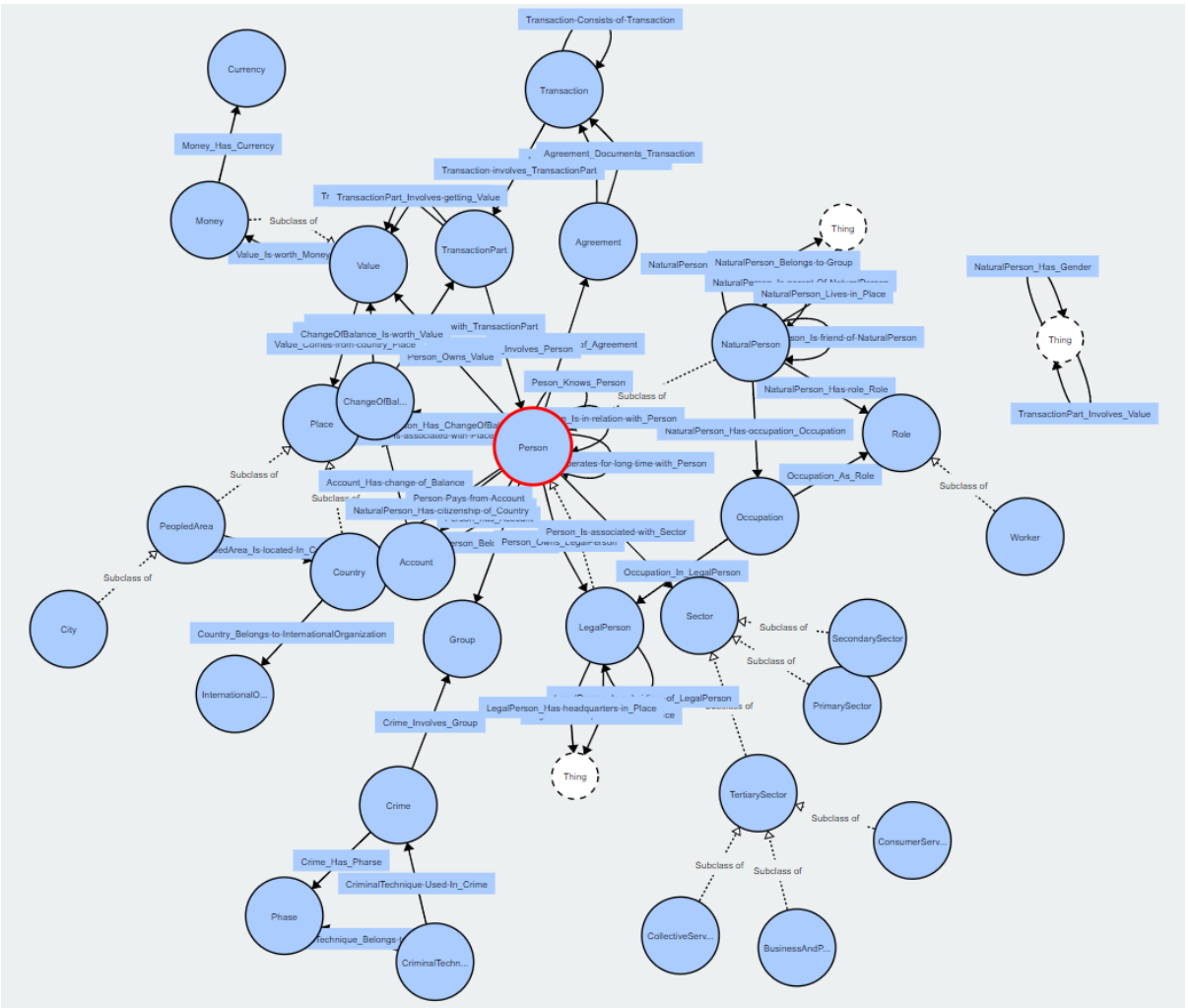


**Fig. 3.** Terminology graph (main concepts and structural relations) defined in Prototype GIIF Ontology - a part of IAFEC ontology modules set– terminology defines rich set of structural relations supporting declarations for reasoning rules.
*Source: Own study.*

In order to detect such frauds, the IAFEC ontology organizes information flow and reasoning paths around a *FinancialCase* and *FinancialTransaction*, which are hubs to which elements of evidence are connected, stating *Organisations*, *Individuals, FinancialChannels, Goods and Services* as well as all available associations organising relations between all involved actors.

The IAFEC analytical mechanisms consist of four consecutive layers (Fig.2):

1. GIIF ontology modules; a universal set of basic terminology formulating financial transactions and their data model;
2. IAFEC ontology modules "surrounding" the GIIF ontology supplementing domain descriptions and delivering additional limitations, ontological rules reflected in the logic (rules or DL constructors);
3. Data transformation mechanisms (services) into instances expressed in IAFEC ontologies;
4. Knowledge processing algorithms and association evaluation algorithms implemented as autonomous components;
5. IAFEC application environment integrating processing components as Protégé 5.0 environment extensions and delivering analytical process reconfiguration.

The division described above allows for the independence of the domain models, and thus for the creation of a set of reusable and reliable base ontologies.

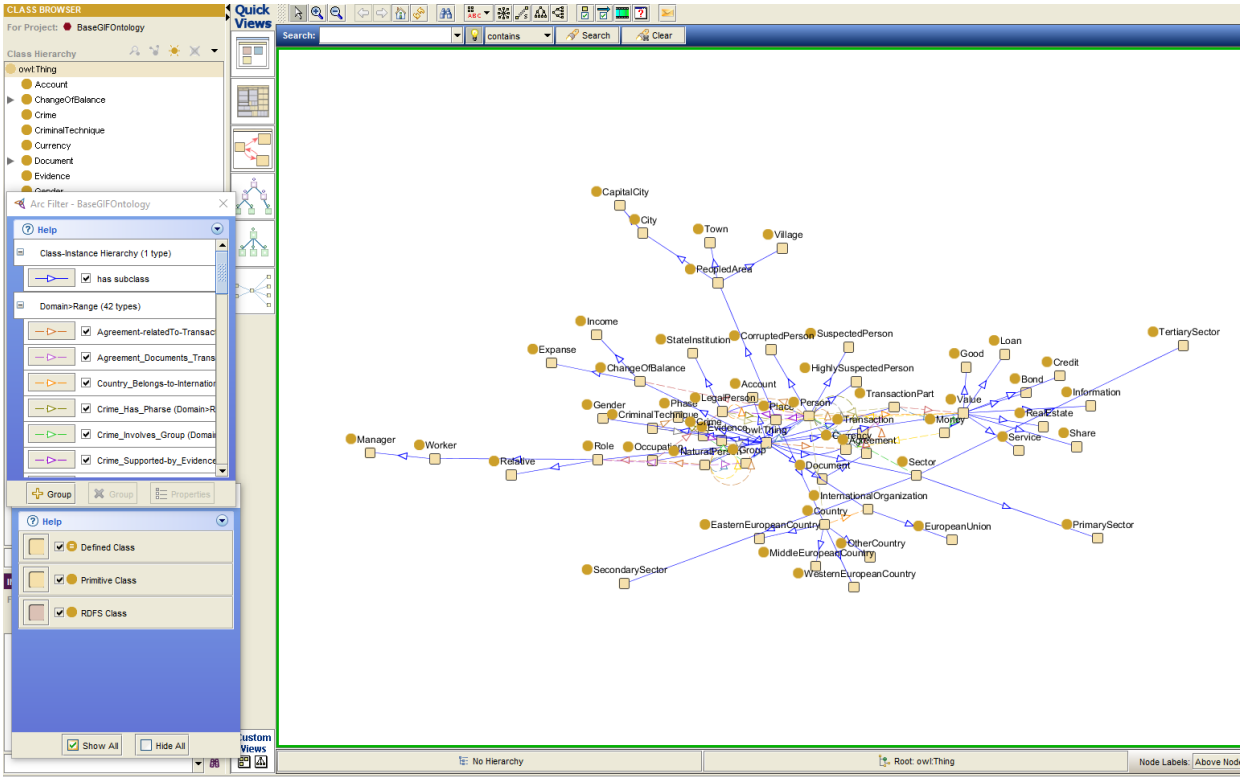


**Fig. 4.** Terminology graph (Protégé – Jambalaya) - main concepts and structural relations) defined in GIIF Financial Data Ontology – financial data record.
*Source: Own study.*

The process of ontology modelling (in developed IAFEC approach) defines ground rules for organizing domain ontologies with respect to terminology and instance data. Ontology modularization helps to cluster and share ontology definitions but most of all supports memory optimization schemes while processing large instance bases.

Developed modelling techniques (CHMIELEWSKI, 2009; CHMIELEWSKI et al., 2016) of ontology modularization have been used in Protégé OWL environment and capture following guidelines:

- cohesive encapsulation of domain concepts in modules, the module's theme should be logically and thematically consistent;
- separate terminology and instance data, therefore move ontology individuals to separate OWL files importing required terminologies (optimize the imports);
- store primitive concepts (OWL classes) and roles defined between primitive concepts (OWL object properties) in separate module, distinct to defined concepts - constructed using restrictions

referencing primitive concepts;
- for ontology integration purposes it is useful to develop separate module storing semantic bridges, definitions of concepts mapping concepts across ontology modules using concept or role equivalence;
- distinction between DL constructors and FOL rules stored in separate modules (BAADER et al. 2003).

Developed IAFEC ontology modules consist of following domain descriptions:
- registry data describing individuals, organisations, companies, vehicles, stakeholders relations, ownership etc. (PESEL, GIIF, CEPiK, CEIDG, KRS).
- social layer description - people, social groups, all interactions with social layer and activities concerned with education, religion, interests, employment, leisure;
- institutions and organizations their description including group members and organization profile and interests;
- financial transactions description – description of atomic financial operations performed by system agents (users, automata) in financial systems, followed by annotation of participants of this process;
- information sources description – a media, sources, content metadata descriptions which enrich the evidence data and provide importance factors.

**Modelling language extensions**

Ontology modules create a layered stack of concepts starting from basic concepts with no logic definitions formulating roles (structural relationships) on top of which, defined concepts are formulated with DL constructs. As a supplement IAFEC ontologies deliver set of SWRL rules, which extend the expressiveness of model:

**Tab. 1.** IAFEC Ontology rules (selected examples) demonstrating the idea of instance classification and recognising important for money loundering (instance) data.

| | |
|---|---|
| Rule-1 | $hiredOnPosition(?\,Person, ?\,Position) \wedge hasPosition(?\,Company, ?\,Position) \wedge BoardMemberPosition(?\,Position) \rightarrow hasBoardMemeber(?\,Person, ?\,Company)$ |
| | $hasCommonBoardMemeberWith$ is transitive. MTA – Monthly Transactions Aggregation |
| Rule-2 | $isSubsidiaryOf(?\,Company1, ?\,Company2) \wedge HighValueMTA(?\,mta) \wedge mtaHasSide(?\,mta, ?\,Company1) \wedge mtaHasSide(?\,mta, ?\,Company2) \rightarrow hasHighValueMTAwithSubs(?\,Company2)$ |
| Rule-3 | $sSubsidiaryOf\,(?\,Company1, ?\,CompanyM) \wedge isSubsidiaryOf\,(?\,Company2, ?\,CompanyM) \wedge HighValueMTA(?\,mta) \wedge mtaHasSide(?\,mta, ?\,Company1) \wedge mtaHasSide(?\,mta, ?\,Company2) \rightarrow hasHighValueMTAbetweenSubs(?\,CompanyM)$ |
| Rule-4 | $ConsideredCompany \equiv (ForeignCompany \sqcap \exists hasHighValueMTAbetweenSubs.\,Company) \sqcup \ldots$ |
| Rule-5 | $mtaHasSide(?\,mta, ?\,Company) \wedge mtaHasSide(?\,mta, ?\,Person) \wedge hasBoardMemeber(?\,Person, ?\,Company) \wedge HighValueMTA(?\,mta) \rightarrow ConsideredMTA(?\,mta)$ |
| Rule-6 | $ForeignTransaction \equiv (Transaction \sqcap \exists transactionHasSide.\,ForeignCompany)$ |
| | $ForeignCompany \equiv \exists hasHeadquatersInCountry.\,\neg Poland$ |

*Source: Own study.*

Presented Tab. 1 contains implementation of example 6 (rules) utilised in an important for reasoning purposes instance insertions into knowledge base. Therefore the method utilizes rules and data aggregation services (presented below) to enrich ontology reasoning capabilities. This feature is performed by delivering means for transforming datatype properties values (data transformation, aggregation, time-dimension recognition) not available in ontology modelling language.

Some characteristics require additional operations to be revealed, e.g. the total value of monthly cash flow between two sides, frequency of transactions, etc. This becomes problematic as description logic supplies datatype-properties but does not provide a means to perform arithmetic or logic operation on their values. Although some SWRL reasoning engines allow that via built-in functions, this language section is out of scope of DL-safe SWRL rules. Even using build-ins to write complex computation formulas in SWRL would seem ill-advised as SWRL was not design for that purpose. In order to solve this problem a java written module was employed. Its task is to gather single transactions-representing individuals with respect to given time period (e.g. month) and its two sides, thus creating a new knowledgebase. This instance shall contain both summary (total transactional value, etc.) and statistical (average/highest transaction value, etc.) values attached via datatype properties. For some analytical purposes it might be required for a summary node to point to its components. Hence, the java module automatically adds object properties to these "source" individuals. Aggregating large volumes of transactions requires high computational-power. For that reason, our suggested approach is to deploy this module on a server and provide a web service as a means of communication with it. The aforementioned module utilizes owl-API library to read and append

to a knowledgebase files and calls HermiT when some minor reasoning tasks are to be performed during the aggregation phase.

**Semantic association formulation and evaluation**

In order to assess importance factor of long sequences, a notion of semantic association (CHMIELEWSKI, 2009; CHMIELEWSKI et al., 2016) could be employed. The semantic association pattern is a chain of intertwined concepts and object-properties. The reasoning engine's goal is to find (in instance base) sequences of individuals and links between them such that they belong respectively to (sub) concepts and (sub)properties defined in the given pattern. The importance of such semantic association (chain) can then be expressed with the help of values obtained from both hierarchical (and in some cases structural) perspective in a manner that promotes nodes and links from lower parts of hierarchy (CHMIELEWSKI, 2009).

Having multi-graph definitions the semantic path definition for ontology structure O and instance base structure IN can be provided, stating interpretations for elements (we consider both: trails (undirected) and path (directed)). Semantic connectivity (concept connectivity, instance connectivity respectively) are sequences node-link permitted by terminology and instance base definitions:

$$S_{con}^{O}\left(c_i, c_j\right) = \left(c_i, r_k, c_{i+1}, \ldots, c_{j-1}, r_l, c_j\right)$$
$$c_i, \ldots, c_j \in C \quad r_k, \ldots, r_l \in R_C$$

$$S_{con}^{IN}\left(x_i, x_j\right) = \left(x_i, y_k, x_{i+1}, \ldots, x_{j-1}, y_l, x_j\right)$$
$$x_i, \ldots, x_j \in I_C^O \quad y_k, \ldots, y_l \in I_{R_C}^O$$

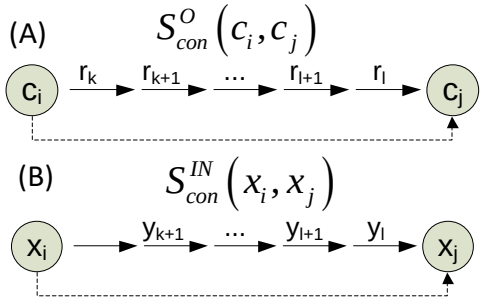


**Fig. 5.** Semantic association definitions in form of semantic connectivity on the level of (A – concepts), (B – instances).
*Source: Own study based on (CHMIELEWSKI, 2011).*

Based on the semantic connectivity and semantic similarity on the data instance level (CHMIELEWSKI, 2009) we can formulate following structures and correspondences which are used by IAFEC environment for association selection and ranking (stating hidden relationships in transactional data).

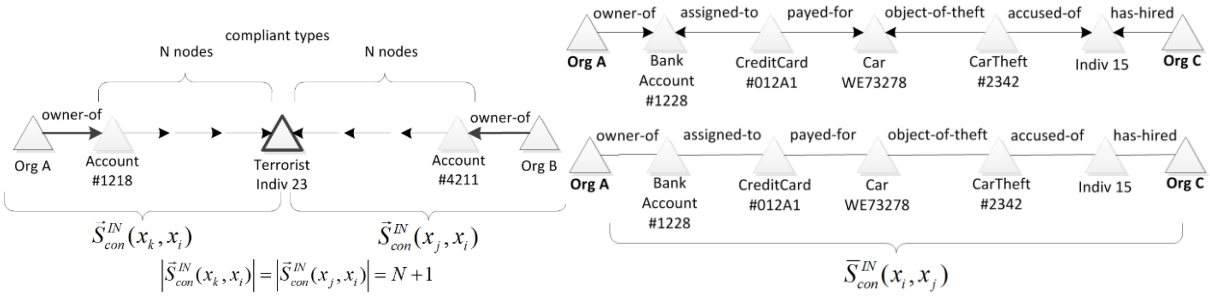


**Fig. 6.** (left) a semantic path (association) demonstrating directed or undirected (relaxed) graph paths in semantic model. (right) example semantic associations in example instance bae demonstrating strict or relaxed connectivity approach applied for IAFEC data.
*Source: Own study based on (CHMIELEWSKI, 2011).*

To demonstrate the structural analysis implemented in IAFEC environment an example semantic association has been presented (Fig.6). The terminology of scenario individuals are described by the IAFEC ontology modules.

The semantic association ranking process utilises measures that form synergic semantic importance score. Conducted analysis helped to understand the nature and trends of selected measures assessing instance data expressed in set of IAFEC terminologies. To prove this assumption, several measures have been chosen to emphasise the semantic importance of particular sets of instances. When analyst combine such measures during analysis he can obtain semantic scores for complex relationships found in knowledge base. The reliability representation of information in the instance base can be obtained as a value correlated with the score of the data source.

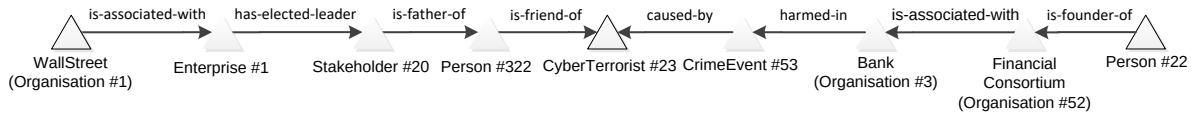


**Fig. 7.** Example semantic association describing hidden context-aware relation between Organization#1 and Person#22 fond in financial transaction data supplemented by the intelligence reports (data about stakeholders and events data).
*Resource: Own study based on (CHMIELEWSKI, 2011).*

**Tab. 2.** Association data instances measures calculated for types (concepts) of instance data formulated in IAFEC ontology modules - measures formulated in (Chmielewski M., 2011) evaluate structural correspondences in constructed and extracted semantic models of financial data.

| **Concept instance** IAFEC Ontologies | CCTM | | | ACCTM | | SCCM (CTG) | | | Concept Score |
|---|---|---|---|---|---|---|---|---|---|
| | CCTM | ACCTM | SCCM (CTG) | ACSD | SCBM | LNCC | ACSD | SCBM | |
| Organisation #1 | 0,2857 | 0,2857 | 0,5000 | 0,5921 | 0,9175 | 0,0476 | 0,6092 | 0,6305 | **0,42237** |
| Enterprise #1 | 0,7143 | 0,7143 | 0,0714 | 0,0263 | 0,0000 | 0,0000 | 0,0345 | 0,0559 | **0,37448** |
| Stakeholder #20 | 1,0000 | 0,4286 | 0,2000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | **043429** |
| Person #322 | 0,3333 | 0,2857 | 0,5000 | 1,0000 | 1,0000 | 0,0778 | 0,9885 | 1,0000 | **0,56225** |
| CyberTerrorist #23 | 1,0000 | 0,7143 | 0,0714 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | 0,0000 | **0,46429** |
| CrimeEvent #53 | 0,7500 | 0,4286 | 0,2000 | 0,0000 | 0,0000 | 0,0000 | 0,0115 | 0,0135 | **0,34862** |
| Organisation #3 | 0,2857 | 0,2857 | 0,5000 | 0,5921 | 0,9175 | 0,0476 | 0,6092 | 0,6305 | **0,42237** |
| Organisation #52 | 0,2857 | 0,2857 | 0,5000 | 0,5921 | 0,9175 | 0,0476 | 0,6092 | 0,6305 | **0,42237** |
| Person #22 | 0,3333 | 0,2857 | 0,5000 | 1,0000 | 1,0000 | 0,0778 | 0,9885 | 1,0000 | **0,56225** |
| **Measure weights** Analyst preference | 35% | 15% | 10% | 15% | 5% | 5% | 10% | 5% | **4,01327** |

*Source: (CHMIELEWSKI, 2011).*

**Tab. 3.** Structural relations measures in given association formulated in IAFEC ontology modules - measures formulated in (Chmielewski M., 2011) evaluate semantic importance of a given relation while expressing data.

| **Role instance (structural relations)** IAFEC Ontologies | ASRD | ASRD in | SRSCTM | ASRCTM | SSRCM | SMSRCM | Relation Score |
|---|---|---|---|---|---|---|---|
| associated-with | 0,0196 | 0,5000 | 0,3333 | 0,1667 | 1,0000 | 1,0000 | 0,44461 |
| has-elected-leader | 0,0000 | 0,5000 | 1,0000 | 0,5000 | 0,1667 | 0,3333 | 0,48334 |
| is-father-of | 0,0000 | 0,5000 | 1,0000 | 0,8333 | 0,0667 | 0,2000 | 0,52167 |
| is-friend-of | 0,0196 | 0,5000 | 0,6667 | 0,3333 | 0,3333 | 0,5000 | 0,41127 |
| caused-by | 0,0196 | 0,5000 | 0,6667 | 0,6667 | 0,1000 | 0,2500 | 0,41796 |
| harmed-in | 0,0000 | 0,5000 | 1,0000 | 0,6667 | 0,1000 | 0,2500 | 0,49834 |
| associated-with | 0,0196 | 0,5000 | 0,3333 | 0,1667 | 1,0000 | 1,0000 | 0,44461 |
| is-founder-of | 0,0000 | 0,5000 | 1,0000 | 0,1667 | 1,0000 | 1,0000 | 0,60834 |
| **Measure weights** Analyst preference | 15% | 15% | 25% | 20% | 15% | 10% | 3,83012 |

*Source: (CHMIELEWSKI, 2011).*

Final aggregated score of selected association is concept instance score = 4,01327, structural relation score = 3,83012 which can be further aggregated using weighting strategy expressing analyst preference model (selecting more important aspects of analysed data – concepts or relations).

## Conclusions

Presented method and its implementation (IAFEC Semantic Association Analyzer), provides a context aware and easily expandable tool for processing semantically relevant financial data that can be used to identify and identify complex frauds. The method delivers ontologies, processing algorithms and measures for evaluating semantics within knowledge bases – making analytical process transparent and verifiable. Selected data instances classified in reasoning process can be explained using reasoner capabilities and further evaluated.

Equipping the method with analytical software, provides declarative possibilities of analytical flow. The method can be integrated with large scale pre-filtering services delivering financial datastreams. In the second stage of processing, the tools described implement human-in-the-loop algorithms that provide configurable and adjustable analytical schemes. The IAFEC environment framework architecture provides an analytically process based workflow that integrates different methods. The presented method provides a complete set of analytical mechanisms for processing, aggregation, filtering and evaluation of data. Network analysis and chart searching are effective tools for specific data processing, therefore the introduction of semantic descriptions and reasoning broadens these possibilities. Scenario analysis and data selection and filtering in a stream are valuable tools for analysts who, after testing the usefulness of the described scheme, can implement the rules for large-scale processing. During the functional testing of the software, a set of fraud data was analysed, providing the expected results of the identification. However, the analyst must take into account the limitations of an ontological processing model based on representation in memory. Recent work on IAFEC semantic tools is moving in the direction of using the Neo4J chart database and triple memory to eliminate memory constraints. In order to support effective management of the ontological model, a modularization methodology has been implemented that simplifies memory management and supports the loading of instance data during the analysis. This supports the knowledge engineer while expanding terminology and provides greater knowledge base processing capabilities. Presented method delivers context-aware and easily extendable tool for processing semantically relevant financial data, which can be used to identify and recognise, complex financial abuses. Encasing the method in form of analytical software such as IAFEC Semantic Association Analyzer tool, delivers declarative analytical flow capabilities, which can be used by analyst to modify and extend terminology or define processing flows. The framework architecture of IAFEC environment provides, a process based analytical workflow, which integrates various methods. Presented method delivers complete set of analytical mechanisms for data transformation, aggregation, filtering and evaluation. Network-based analysis, graph querying are effective tools for specific data processing, thus introduction of semantic descriptions and reasoning extend these capabilities. Scenario-based analysis as well as in-stream data selection and filtering are valuable tools for analysts which after testing the usefulness of described scheme can deploy the rules for large scale processing. IAFEC Semantic Association Analyzer can be integrated with preliminary filtering services of large scale data provided by financial institutions and aggregated by auditing institutions such as GIIF. In the second stage of processing, the described tools implement human-in-the-loop algorithms, which deliver configurable and adjustable analytical schemes. Provided architectural solutions deliver capabilities for seamless integration within larger heterogeneous SOA environment. During software functional tests, a set of financial fraud datasets have been verified and analysed, providing expected identification results in the limited size financial transfers database. However, the analyst needs to consider the limitations of ontology processing model relying on in-memory representation. The complexity of reasoning mechanisms and memory footprint of Ontology models require additional effort to apply the method in data stream processing, therefore developed environment integrates graph databases and triple stores.

## Acknowledgements

This work was partially supported by research projects DOBR/0073/R/ID1/2012/03: "Advanced ICT techniques supporting data analysis processes in the domain of financial frauds" and DOBR/0023/R/ID3/2013/03 supported by the National Center For Research and Development.

## References

BAADER, F., MCGUINNESS, D., NARDI, D., PATEL-SCHNEIDER, P. 2007. *The Description Logic Handbook: Theory, implementation, and applications*. Cambridge University Press.

BARTHELEMY, M., CHOW, E., ELIASSI-RAD, T. 2005. *Knowledge Representation Issues in Semantic Graphs for Relationship Detection*. In: AAAI Spring Symposium: AI Technologies for Homeland Security.

BASILI, R., CAMMISA, M., PENNACCHIOTTI, M., ZANZOTTO, F. M. 2003. *Ontology-driven Information Retrieval in FF-Poirot*.

CHMIELEWSKI, M. 2009. *Ontology Applications for Achieving Situation Awareness in Military Decision Support Systems.* ICCCI 2009, LNCS, 5796: 528-539.

CHMIELEWSKI, M., STĄPOR, P. 2011. *Protégé based environment for DL knowledge base structural analysis*. In: Computational Collective Intelligence. Technologies and Applications, p. 314-325.

CHMIELEWSKI, M. 2011. *Ontology-based association assessment method using graph and logic reasoning techniques.* Military University of Technology, Warsaw.

CHMIELEWSKI, M., STĄPOR, P. 2016. *Medical Data Unification Using Ontology-Based Semantic Model Structural Analysis*. In: Information Systems Architecture and Technology: Proceedings of 36th International Conference on Information Systems Architecture and Technology - ISAT 2015.

DENTLER, K., CORNET, R., TEN TEIJE, A. DE KEIZER, N. 2011. *Comparison of Reasoners for large Ontologies in the OWL 2 EL Profile*. In: Semantic Web, 2.

GRUBER, T. R. 1993. *Toward principles for the design of ontologies used for knowledge sharing*. In: Formal Ontology in Conceptual Analysis and Knowledge Representation, Kluwer Academic (in preparation).

PMoF, 2012. Poland's Ministry of Finance, Department of Financial Information, http://www.mf.gov.pl/ministerstwo-finansow/dzialalnosc/giif/system (access 01.04.2018).

PROTÉGÉ WEB, 2016. Protégé OWL wiki webpage, https://protegewiki.stanford.edu/wiki/Main_Page (access: 01.04.2018).

JARRAR, M., MEERSMAN, R. 2008. *Ontology Engineering, The DOGMA Approach*. In: Advances in Web Semantics I. LNCS, 4891, Springer.

STAAB, S., STUDER, R. 2004. *Handbook on Ontologies*. Springer.