Protégé Based Environment for DL Knowledge Base Structural Analysis

Mariusz Chmielewski and Piotr Stąpor

Computer Science Department,
Cybernetics Faculty, Military University of Technology,
Kaliskiego 2, 00-908 Warsaw, Poland

Abstract. Structural analysis of knowledge bases improves process of obtaining additional hidden information and provides new means of information discovery. In this paper we propose a method of structural analysis and characterise in details developed toolkit as a Protégé extension. The aim of method itself is to identify hidden relationships using structural weighted graph analysis, applied for both terminology and instance base. In order to evaluate new relationships between instances, we have chosen a set of measures that consider the analysis of both vertices and links, thus providing new semantic information for identified associations. Developed environment concludes both approaches (logic based and graph based) in order to apply reasoning techniques, graph based algorithms and measures for inferring important analytical information. The scope of paper provides also application areas of such method in terms of data mining techniques for crisis identification and multi-criteria data analysis.

Keywords: semantic association, semantic similarity, ontology, OWL, description knowledge.

1 Introduction

Knowledge bases exploiting the Semantic Web [1] idea are becoming more and more competitive to legacy databases in terms of network data processing and analytical capabilities. The utilization of semantic networks as a mean of data representation, fused with the power of Description Logic [2] and reasoning mechanisms introduces enhancements in the field of automatic information discovery. Already available DL reasoners (ex. Pellet, Keon2, HermiT) and rule-based engines (ex. Jess) expose the abilities of instance classification process based on consistent rules and model validation. Protégé OWL [13], among many modelling features, offers a set of analytical functions such as ontology metrics and model visualisation, but at the same time it lacks algorithms for structural KB analysis. This gap can be filled by described Terrana plug-in, which concentrates mainly on applying ontology and instance base evaluation techniques for hidden information discovery. Term hidden information corresponds to all indirect relations between concepts or instances added after process of multi-criteria semantic association analysis. Evaluation of model characteristics is

P. Jędrzejowicz et al. (Eds.): ICCCI 2011, Part I, LNCS 6922, pp. 314–325, 2011.

performed on weighted multi-graph representation. Terrana implements model transformation rules to create multi-graph which reflects ontology or instance base elements connected with structural relations. Developed model is further used as a basis for evaluating characteristics of nodes and links considering both taxonomical and axiomatic information. Detailed information on transformation process and its accuracy and will be discussed later.

It is crucial to state that multi-criteria ranking process is achieved by introducing relevant ontology (concept, relationship) and instance base metrics, which provide evaluation of a wide spectrum of characteristics. Sources [9][12][11] distinguish variety of graph based factors which concentrate mainly on structure assessment, which we consider but also modify to drain semantics. Algorithms concentrate on element importance depending on centrality measures, such as: degree, connectivity, diameter etc. The ability to find hidden associations, combined with a set of algorithms for asserted relations linkage and evaluation is essential.

2 Semantic Model Formal Definition

In order to understand ranking process we should lay out the theoretical model on which method and software rely. We start with altered definition of ontology [6] formulated as:

$$O = \left\langle C, R_C, R_R^H, R_R^I, A^O \right\rangle \quad \text{(Eq.1)}$$

where C is a set of identified unique concepts, R_C – is a set of relations between defined concepts and A^O - is a set of axioms defined for ontology O. Set $R_C = H_C \cup S_C$ is additionally divided based on the characteristics of relations to set S_C structural relations and hierarchical H_C relations used to organise concept taxonomy. R_R^H is a set of relations between elements of R_C identifying relation taxonomy and R_R^I is a set of relations between elements of R_C identifying specific semantics for chosen relations used to express inverse relations.

Using presented ontology O definition, there can be formed an instance base (data level knowledge base which is used to store the instances of elements defined on the conceptual level), defined as:

$$IN^O = \left\langle I_C^O, I_{R_C}^O, V_C^O, V_{R_C}^O \right\rangle \quad \text{(Eq.2)}$$

, where: I_C^O contains instances of concepts C, I_R^O contains instances of relations R_C in a given ontology O. Elements of instance base can be defined as follows:

$I_C^O = \bigcup_{c \in C} Inst_C^O(c)$, $Inst_C^O(c)$, identifies set of all instances for given concept $c \in C$;

$I_{R_C}^O = \bigcup_{r \in R_C} Inst_{R_C}^O(r)$, identifies set of all instances of relation $r \in R_C$ which meet

$Inst_{R_C}^O(r) = \left\{ (x_i, x_j) \in I_C^O \times I_C^O : r = \left(V_C^O(x_i), V_C^O(x_j) \right) \wedge r \in R_C \right\}$;

$V_C^O : I_C^O \to 2^C$ is a classifier function which reflects possible types of instances I_C^O;
$V_{R_C}^O : I_{R_C}^O \to 2^{R_C}$ is a relations classifier function, identifying a set of relation types for a chosen relation instance $I_{R_C}^O$;

The term semantic model shall, in further part of this paper, be understood as a pair:

$$M^{Sem} = \left(O, IN^O\right) \quad \text{(Eq.3)}$$

Ontology structure for further structural analysis will be defined as weighted multi-graph:

$$\overline{\overline{O}} = \left\langle \Omega, \left\{ f_k\left(c\right)\right\}_{k\in\{1,\dots,LF\},c\in C}, \left\{ g_l\left(r\right)\right\}_{l\in\{1,\dots,LG\},r\in R_C} \right\rangle \quad \text{(Eq.4)}$$

where Ω is ontology graph structure on which we define families of functions for nodes-concepts $\left\{ f_k\left(c\right)\right\}_{k\in\{1,\dots,LF\},c\in C}$ and links-relations $\left\{ g_l\left(r\right)\right\}_{l\in\{1,\dots,LG\},r\in R_C}$.

Multigraph $\Omega = \left\langle C, R_C, P\right\rangle$, contains nodes C, edges R_C and triple relation $P = \left\{\left\langle c_i, r, c_j\right\rangle : c_i, c_j \in C \wedge r \in R_C\right\}$, fulfilling conditions:

$$\forall_{r\in R_C} \exists_{\langle c_i,c_j\rangle \in C\times C} \left\langle c_i, r, c_j\right\rangle \in P \qquad \forall_{r\in R_C} \forall_{\substack{c_i,c_j\in C \\ c_k,c_l\in C}} \left\{ \begin{array}{l} \left[\left\langle c_i, r, c_j\right\rangle \in P \wedge \left\langle c_k, r, c_l\right\rangle \in P\right] \Rightarrow \\ \left[\left(c_i = c_k\right)\wedge\left(c_j = c_l\right) \vee \left(c_i = c_l\right)\wedge\left(c_j = c_k\right)\right] \end{array} \right\}$$

Weighted multi-graph elements stand for:
$f_k : C \to val_k^C$ is the k-th function described on the multi-graph's nodes (concepts), $k = 1,\dots,LF$, (LF – number of f functions); val_k^C is a k-th set of values describing concepts, $g_l : R_C \to val_l^{R_C}$ – the l-th function described on the multi-graph's links (relations), $l = 1,\dots,LG$ (LH–number of h functions), $val_l^{R_C}$ is a l-th set of values describing relations.

Using previous definitions we can formulate an instance network structure model which is used to store the instances of ontology elements, providing data level description:

$$\overline{\overline{IN^O}} = \left\langle \Phi, \left\{ \varphi_h\left(x\right)\right\}_{h\in\{1,\dots,LH\},x\in I_C^O}, \left\{ \phi_l\left(y\right)\right\}_{l\in\{1,\dots,LL\},y\in I_{R_C}^O} \right\rangle \quad \text{(Eq.5)}$$

where Φ is an instance base structure multi-graph $\Phi = \left\langle I_C^O, I_{R_C}^O, \overline{P}\right\rangle$, which contains nodes I_C^O, edges $I_{R_C}^O$ (interpretation same as IN^O) and triple relation defined as $\overline{P} = \left\{\left\langle x_i, y, x_j\right\rangle : x_i, x_j \in I_C^O \wedge r \in I_{R_C}^O\right\}$, fulfilling conditions:

$$\forall_{y\in I_{R_C}^O} \exists_{\langle x_i,x_j\rangle \in I_C^O\times I_C^O} \left\langle x_i, y, x_j\right\rangle \in \overline{P} \qquad \forall_{y\in I_{R_C}^O} \forall_{\substack{x_i,x_j\in I_C^O \\ x_k,x_l\in I_C^O}} \left\{ \begin{array}{l} \left[\left\langle x_i, y, x_j\right\rangle \in \overline{P} \wedge \left\langle x_k, y, x_l\right\rangle \in \overline{P}\right] \Rightarrow \\ \left[\left(x_i = x_k\right)\wedge\left(x_j = x_l\right) \vee \left(x_i = x_l\right)\wedge\left(x_j = x_k\right)\right] \end{array} \right\}$$

on which we define families of functions for nodes-concept instances $\left\{\varphi_h\left(x\right)\right\}_{h\in\{1,\dots,LH\},x\in I_C^O}$, edges-relation instances $\left\{\phi_l\left(y\right)\right\}_{l\in\{1,\dots,LL\},y\in I_{R_C}^O}$. Respectively $\varphi_h : I_C^O \to val_h^{I_C^O}$ is the h-th function described on the multi-graph's nodes (instances of concepts), $h = 1,...,LH$ (LH – number of φ functions), $val_h^{I_C^O}$ - is a h-th set of values describing concept instances, $\phi_l : I_{R_C}^O \to val_l^{I_{R_C}^O}$ is the l-th function described on the multi-graph's links (instances of relations), $l = 1,...,LM$ (LM–number of ϕ functions), $val_l^{I_{R_C}^O}$ is a l-th set of values describing relation instances.

Having multi-graph definitions we can describe semantic paths definitions for ontology structure $\overline{\overline{O}}$ and instance base structure $\overline{\overline{IN^O}}$ and provide interpretations for those elements (we consider both: trails (undirected) and path (directed)). Semantic connectivity (concept connectivity, instance connectivity respectively) are graph like chains (undirected) or paths (directed) permitted by both ontology and instance base definitions:

$$S_{con}^{O}\left(c_i, c_j\right) = \left(c_i, r_k, c_{i+1}, ..., c_{j-1}, r_l, c_j\right),\ c_i, ..., c_j \in C,\ r_k, ..., r_l \in R_C$$
$$S_{con}^{IN}\left(x_i, x_j\right) = \left(x_i, y_k, x_{i+1}, ..., x_{j-1}, y_l, x_j\right),\ x_i, ..., x_j \in I_C^O,\ y_k, ..., y_l \in I_{R_C}^O.$$

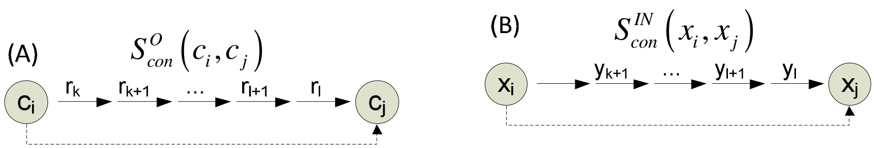


Fig. 1. Semantic connectivity (A) for concepts, (B) for instances

3 Terminology and Instance Base Measures

Some valuable information can be obtained by studying the ontology or instance base multi-graphs depending on proper vertices and links measure definitions. We base our assumptions on several observations:

- semantic model consist of taxonomical, structural and axiomatic information;
- each of those aspects can be evaluated using consistent approach which mainly require custom build structure (similar to presented earlier);
- ranking method must consider multi-criteria evaluation due to emphasised analysis approach;

As an example we can call in $Rank_{st}^{O}\left(c\right) = \frac{1}{k^{\deg(c)}}, c \in C, k > 1$, a degree of a vertex determines the informational usefulness of an ontology concept depending on the connection popularity of such node. In order to find and estimate the information value of the semantic associations, Terrana plug-in supplies user with following ranking factors: Semantic Link Relevance (SLR), Concept Clustering Coefficient (CCC), Concept Taxonomical Depth Measure (CTDM), and Semantic Concept Connectivity Measure (SCCM). Further part of this paper exploits in detail the aim of above mentioned measures and their usage in the field of hidden relation analysis.

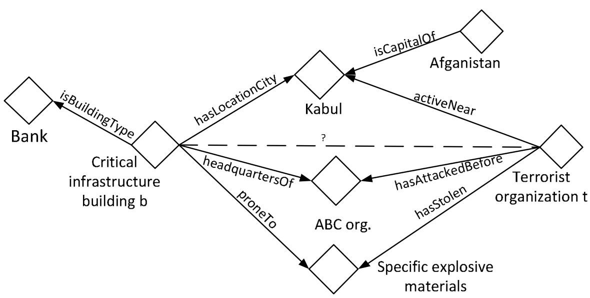
Semantic Link Relevance has been introduced in the [9] and serves the purpose of determining the importance of existent or nonexistent link between two individuals or concepts. While small SLR value may indicate, that the link has little importance (and thus should be removed from semantic model), the large enough value may be a reason for “materializing”, the nonexistent (*potential* in [9]) link. The SLR parameter can also be used to make the semantic model analysis process more effective, by pointing the links in ontology or instance base multi-graphs that require more attention. SLR is expressed as the ratio of common neighbors to all neighbors owned by a pair of either individuals or concepts. Based on definitions of Link Relevance [9] we redefine measure in respect of ontology and instance base structures:

$$S_{rel}^{O}\left(c_i,c_j\right)=\frac{\left|N\left(c_i,c_j\right)\right|}{\left|T\left(c_i,c_j\right)\right|}\wedge c_i,c_j\in C\text{'}\quad S_{rel}^{IN}\left(x_i,x_j\right)=\frac{\left|N\left(x_i,x_j\right)\right|}{\left|T\left(x_i,x_j\right)\right|}\wedge x_i,x_j\in I_C$$

where, having regard to presented multi-graph model definitions on generic level:

$$S_{rel}\left(v_i,v_j\right)=\frac{\left|N\left(v_i,v_j\right)\right|}{\left|T\left(v_i,v_j\right)\right|}\qquad \begin{array}{l} N\left(v_i,v_j\right)=\left\{v_k\mid v_k \text{ is linked to } v_i\wedge v_j, v_k\neq v_i, v_k\neq v_j\right\}\\ T\left(v_i,v_j\right)=\left\{v_k\mid v_k \text{ is linked to } v_i\vee v_j, v_k\neq v_i, v_k\neq v_j\right\}\end{array}$$

Example 1. Calculating Semantic Link Relevance for (instance base)



$where\ multigraph\ vertices\ v\in I_C^0\ \wedge e\in I_{R_C}^0$

$b = Critical\ infrastructure\ building\ b,\qquad t = terrorist\ organization\ t$

$N(b,t) = \{Kabul, ARB\ org., explosive\ materials\}$

$T(b,t) = \{\{bank\}\cup N(a,b)\}$

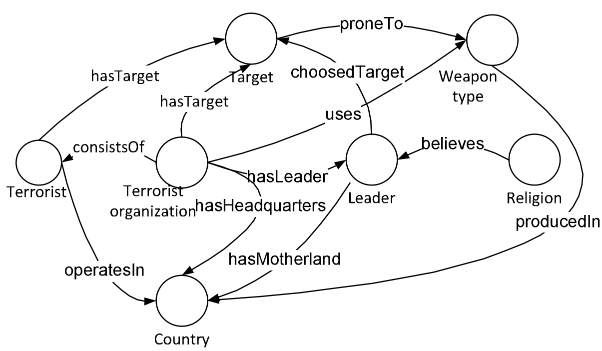
$$S_{rel}^{IN}\left(b,t\right)=\frac{\left|N\left(b,t\right)\right|}{\left|T\left(b,t\right)\right|}=\frac{3}{4}=0.75$$

High Semantic Link Relevance measure indicates existence of asserted linkage between two given instances: critical financial infrastructure bank - building *b* and terrorist organization *t*. Such measure can be used to identify if additional relation should be introduced to link entities on terminological level (concepts).

Considering the usefulness of SLR measure, there is also a need of defining a method of vertex importance computation. For this purpose, Terrana implements, Concept Clustering Coefficent (further referred as CCC) evaluation algorithm. In [9], authors define CCC for an ontology with the help of formula: $CCC(c_i) = \frac{E(c_i)}{k_c(c_i)\left[k_c(c_i)-1\right]/2}$, where $c_i \in C$ in the defined model, $E(c_i)$ represents links quantity between the nearest neighbors of vertex c_i and $k_c(c_i) = \deg(c_i)$. In other words, the CCC value of a certain node c_i can be expressed as the ratio of number of existing connections between c_i neighbors divided by the quantity of all allowable connections between c_i neighbors.

The representation of ontology as a multi-graph implies the slight $CCC(c_i)$ adjustment which relays on large number of possible links existence. Algorithm treats any connections quantity between two neighbors with the same direction as one (as a result, the maximum number of links, taken into account during CCC evaluation, between any pair of vertexes is two). Additionally, because the model used in Terrana is directed multi-graph, the removal of division by two in denominator is implied. Alternatively it is possible to solve this issue, by constructing a multi-graph's skeleton first, and then utilizing it in the process of CCC evaluation. Based on [9] we may assume that CCC greater than certain threshold, may be used to evaluate importance of node in terms of concept connectedness allowing us to infer that concept is worth considering in further analysis.

Example 2. Calculating Concept Clustering Coefficient example



to = Terrorist Organization, wt = WeaponType

$Neigh(c_i)$ – neighbors of given c_i *concept*

Neigh(to) = {Terrorist group, Target, Leader, Country, wt }

$$k_{to} = |Neigh(to)| = 5$$

Existing connections between *to* neighbors *{hasTarget, operatesIn, hasMotherland, choosesTarget, hasTarget, proneTo, producedIn}*

$\because k_{to}(k_{to} - 1) = 20 \wedge E_{to} = 7 \;\therefore\; CCC(to) = 0.35$

$\because k_{wt}(k_{wt} - 1) = 6 \wedge E_{wt} = 2 \;\therefore\; CCC(to) = 0.33$

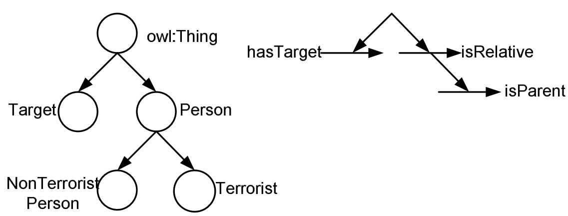
From the above, we may conclude, that for provided semantic model, the knowledge of used weapon type is less useful, than information about the terrorist organization which uses it.

3.1 Taxonomical Measures

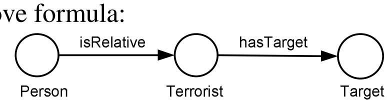
The taxonomical measures rely on ontology hierarchical relations providing the concepts' information load, stored in form of generalization-specialization tree. In Protégé, the taxonomy is stretched between owl:Thing (domain) and owl:Nothing (empty set) concepts. The deeper the placement within the taxonomy, the more detailed concept is (thus carrying more valuable semantic information).

Semantic association evaluation method utilizes the mentioned earlier hierarchy property. The informational level value for every concept (vertex) or relation (link), can be calculated using Concept Taxonomical Depth Measure (CTDM) formula: $I_{val}(c) = \frac{lvl(c)}{H(c)}$, where $lvl(c)$ function returns the level of c element in defined taxonomy, and $H(c)$ is the number of levels of whole hierarchy. For either classes or properties instances, we can determine their usefulness by evaluating $I_{val}(V_C^O(x))$ or $I_{val}(V_{R_C}^O(y))$ respectively ($x \in I_C^O, y \in I_{R_C}^O$). Those functions classify the instance thus helping to incorporate taxonomical measures for given instance. Having evaluated values for model elements, it is possible to estimate a given semantic association, expressed as simple acyclic graph path, as a following product $\prod_{a \in A_e} I_{val}(a)$, where A_e is a set of vertexes and links belonging to investigated association.

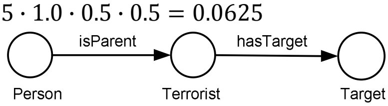
Example 3. Definition of concepts and relations taxonomies



In order to estimate informational value of two semantic associations, we will use the presented above formula:



$\prod_{a \in A_e} I_{val}(a) = I_{val}(Person) \cdot I_{val}(isRelative) \cdot I_{val}(Terrorist) \cdot I_{val}(hasTarget) \cdot I_{val}(Target) = 0.5 \cdot 0.5 \cdot 1.0 \cdot 0.5 \cdot 0.5 = 0.0625$



$\prod_{a \in A_e} I_{val}(a) = 0.125$

As shown, the more precise (specific) are the elements of studied association, the more useful they are. The information about terrorist's parent is more valuable than one about his (perhaps distant) relative.

3.2 Semantic Connectivity Based Measures

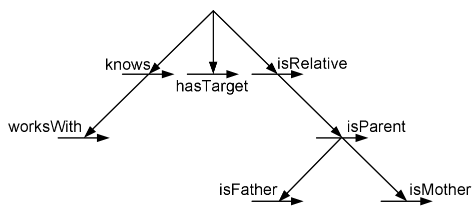
In order to improve Terrana's semantic model capabilities, we have chosen additional taxonomical measures [11], which rely on aggregated semantic connectivity S_{con}^O weighted measure:

- **Concept rank** $t_C^{rank}(c_i) = \frac{1}{h^{depth^C(c_i)}}$, where h > 1 and $depth^C(c_i) = lvl(c_i)$ returns the depth of a concept c_i in ontology concepts hierarchy tree.
- **Concept instance rank** $t_{I_C^O}^{rank}(x_k) = t_C^{rank}(V_C^O(x_k))$.
- **Relationship rank** $t_R^{rank}(c_i, c_j) = \frac{1}{h^{depth^R(c_i,c_j)}}$, h > 1. $depth^R(c_i, c_j) = lvl(r(c_i, c_j) \in R_C)$ returns the depth of a property in an ontology properties hierarchy tree.
- **Direct relationship rank** $t_{I_{R_C}^O}^{rank}(x_k, x_l) = t_R^{rank}(V_C^O(x_k), V_C^O(x_l))$

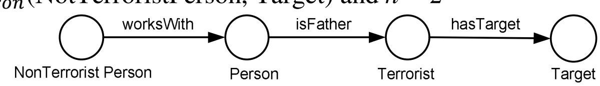
Considering above definitions, the final way of ranking given semantic association is as follows [11][9]:

$$t_{con}^{rank}(c_m, c_n) = t_{con(C)}^{rank}(c_m, c_n) \times t_{con(R)}^{rank}(c_m, c_n) \quad \text{where:}$$
$$t_{con(C)}^{rank}(c_m, c_n) = \prod_{c' \in S_{con}^O(c_m, c_n)} t_C^{rank}(c')$$
$$t_{con(R)}^{rank}(c_m, c_n) = \prod_{(c', c") \in S_{con}^O(c_m, c_n)} t_R^{rank}(c', c")$$

Example 4. Estimating semantic association



Let s = S_{con}^O(NotTerroristPerson, Target) and $h = 2$



$$t_{con(C)}^{rank}(NotTerroristPerson, Target) = 0.25 \cdot 0.5 \cdot 0.25 \cdot 0.5 = 0.01562$$
$$t_{con(R)}^{rank}(NotTerroristPerson, Target) = 0.25 \cdot 0.125 \cdot 0.5 \cdot 1 = 0.01562$$
$$t_{con}^{rank}(NotTerroristPerson, Target) = 0.01562 \cdot 0.01562 \approx\ 0{,}00024$$

4 Multi-criteria Association Evaluation

Depending on the analyst requirements, presented method is able to provide adjustments for evaluating structural measures and ranking semantic associations.

Each measure is connected with a certain evaluation criteria. For example: $m_{hier}^c = \frac{lvl(c)}{h^{depth^C(c)}}, where\ c \in C$ uses specialization level as its criterion. $m_{dist}^{c_i,c_j}$ would promote classes less edges apart from certain important concept c_j, and m_{deg}^cwould judge the importance basing on number of neighbours of a given c vertex (degree). The problem of choosing the right value arises - how to interpret the set of results from different methods. To solve this problem, we recommend the use of multi-criteria approach:

- define measures m_i $i \in \overline{1,n} = \{1,2,3\ldots n\}$, where n$\in \mathbb{N}$, additionally for each measure $m_i \in \langle 0,1 \rangle$
- define weights w_i $i \in \overline{1,n}$, such that $\sum_{i=1}^{n} w_i = 1$;
- define vector M, such that $M^T = [m_1 \quad m_2 \quad \cdots \quad m_n]$;
- define vector W, such that $W^T = [w_1 \quad w_2 \quad \cdots \quad w_n]$;
- count the aggregated measure $M_{agg} = W \cdot M = W^T M = \sum_{i=1}^{n} w_i m_i$

Using prepared evaluation tools presented in this work so far, it is possible already to appraise usefulness of: concepts, individuals, properties, instances of properties, and in the end semantic associations. Additionally, environment provides information concerning the whole structure of the multi-graph. The summary for testbed instance base has been placed beneath.

Table 1. Instance data evaluation based on proposed graph based measures ranking importance on instance base nodes (facts) and their semantic importance inside KB

Order:65 Size: 55	Degree \in,out	CCC	Betweenness	Closeness	Eigenvector Centrality	PageRank (0.15)	HITS hub (0.5)	HITS (0.5)
September11Attacks	0 /8,8	0.0	0	1.7619	0.0123	0.0129	0.1184	0.0060
HajjKhalilBanna	1 /4,5	1.0	0	0.8000	0.0153	0.0151	0.4348	0.1763
PlaneHijack	1 /0,1	0.0	0	0.0000	0.1377	0.0143	0.0161	0.0364
SuicideAttacks	1 /0,1	0.0	0	0.0000	0.0137	0.0143	0.1607	0.0364
MilitaryBuildings	0 /0,0	0.0	0	0.0000	0.0122	0.0129	0.1607	0.0060
AbuNidalOrg.	2 /3,5	0.3	10	1.5454	0.0459	0.0368	0.0781	0.1928
AlQaeda	2 /7,9	0.0	18	1.3077	0.0162	0.0164	0.0540	0.1555
OsamaBinLaden	1 /6,7	0.0	10	1.3846	0.0146	0.0149	0.4642	0.0199
Hezbollah	1 /6,7	0.0	11	0.8571	0.0151	0.0154	0.0510	0.0904
PalestineLiberationOrg.	2 /1,3	1.0	0	2.3636	0.0306	0.0255	0.0642	0.1964

Link Name	*hasLeader*	*hasGender*	*hasConnectionWithTerroristOrg*	*hasEthnicity*
CLR	0.1429	0.1667	0.5000	0.1429

5 Environment Architecture

We have chosen to implement all mechanisms of association discovery in Protégé due to its modelling and ontology processing capabilities. This approach has been motivated to maximize reuse of already developed software components such as: knowledge base operations, inference mechanisms integration, modelling language processing etc. Plugin capabilities delivers different semantic model usage compared to those directly found in Protégé OWL, which concentrates mainly on logical and consistent model design.

Review of those capacities confronted with association analysis method needs, helped us to identify requirements for future extensions addressing:

- weighted graph transformation mechanisms;
- domain terminology evaluation (classes, properties);
- instance base data assessment (individuals) and propagation of found premises to ontology model;
- complex path evaluation for terminology and instance base;

Reuse of available components concentrated on using Jess SWRL rule processing engine and Pellet as DL reasoner. The solution relies heavily on graph processing framework provided by JUNG library – used for ontology and instance base model representations, their visualization and invariants evaluation.

Model transformer component scans the structure of supplied by Protégé knowledge base and maps it to a JUNG based multi-graph according to prepared rules. Further processing includes graph invariants computation using JUNG library supplied algorithms (ex. Betweenness, Closeness, Eigenvector Centrality, PageRank, HITS) and our internal implementation of (Concept Clustering Coefficient, Concept Link Relevance).

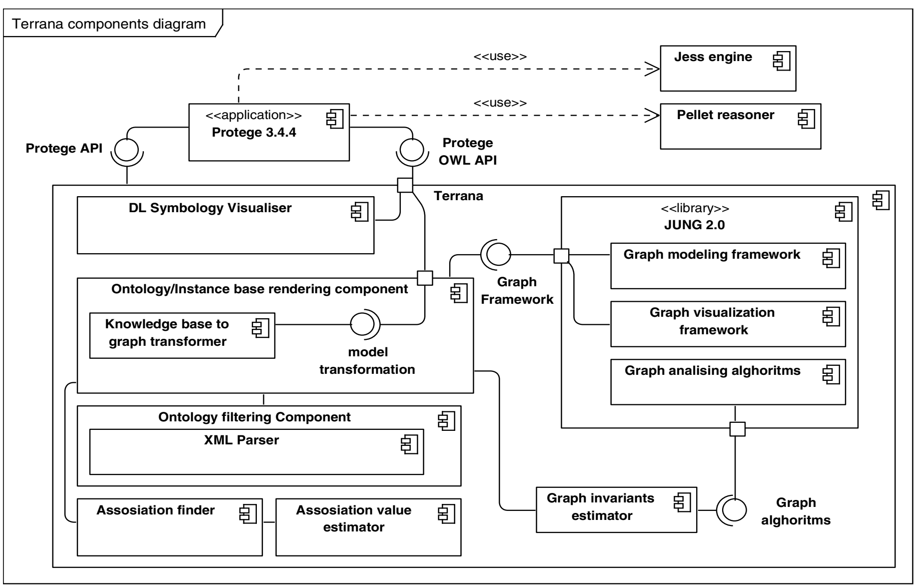


Fig. 2. Developed environment components and used external libraries

Multi-graph representations are utilized by semantic connectivity algorithm which concentrates on searching for a simple acyclic path (or all such paths) between given two concepts or individuals. The association value estimator computes the importance/informational value of such discovered connection using a selected evaluation method.

6 Summary

Information gathering in knowledge base may introduce new approach in analysis while applying flexible graph structures. Presented approach lay foundations and exploits many possible applications for knowledge base validation and data mining techniques. Development of new, more effective ontology measures is followed by a research explaining how those measures can assess instant data, identify hidden relationships and evaluate terminology design. Implemented method concentrates on distinct quantitative analysis of ontologies and instance bases for information discovery and validation. Method utilises multi-criteria expert tuned approach in order to extend structural KB elements evaluation based on known node importance measures applied on the ground of semantic models.

Designed environment has been applied for solving terrorism related problems using adaptive information gathering and to semantic model representation mapping techniques. Processing of such knowledge base may be performed by logical reasoning and structural analysis, which mainly concentrates on: (a) finding hidden transitive relations between chosen instances; (b) classification of concept instances based on gathered characteristics and defined rules, (c) prediction of possible events (based on structure of certain vertexes and relations between them); (d) modeling complex relationships between main instances.

References

1. Davies, J., Studer, R., Warren, P.: Semantic Web Technologies: Trends and Research in Ontology-based Systems. John Wiley & Sons, Chichester (2006)
2. Baader, F., McGuinness, D., Nardi, D., Patel-Schneider, P.: The Description Logic Handbook: Theory, implementation, and applications. Cambridge University Press, Cambridge (2007) ISBN 0521876257
3. Cardoso, J.: Semantic Web Services: Theory, Tools and Applications, . IGI Global (March 2007) ISBN-10: 159904045X
4. Segaran, T.: Programming Collective Intelligence: Building Smart Web 2.0 Applications. O'Reilly Media, Sebastopol (2007) ISBN 0596529325
5. Staab, S., Studer, R.: Handbook on Ontologies. Springer, Heidelberg (2004) ISBN 3540408347
6. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. In: Formal Ontology in Conceptual Analysis and Knowledge Representation. Kluwer Academic, Boston (1993) (in preparation)
7. Gruber, T.R., Olsen, G.R.: An Ontology for Engineering Mathematics
8. Chmielewski, M., Gałka, A., Jarema, P., Krasowski, K., Kosiński, A.: Semantic Knowledge Representation in Terrorist Threat Analysis for Crisis Management Systems. LNCS. Springer, Heidelberg (2009)
9. Barthelemy, M., Chow, E., Eliassi-Rad, T.: Knowledge Representation Issues in Semantic Graphs for Relationship Detection, UCRL-CONF-209845

10. Chmielewski, M., Kasprzyk, R.: Multi-criteria semantic association ranking based on instance knowledgebase analysis for criminal organisation identification. In: Euro 2010 (2010)
11. Ge, J., Qiu, Y.: Concept Similarity Matching Based on Semantic Distance. Faculty of Computer and Information Science. Southwest University Chongqing, China
12. Baur, M., Benkert, M.: Network comparison. In: Brandes, U., Erlebach, T. (eds.) Network Analysis. LNCS, vol. 3418, pp. 318–340. Springer, Heidelberg (2005)
13. Protégé OWL webpage, `http://protege.stanford.edu/overview/protege-owl.html`