Probabilistic graphical models for multi-source fusion from text sources

Georgiy Levchuk
Aptima Inc.
12 Gill Street, Suite 1400, Woburn, MA, USA 01801
georgiy@aptima.com

Erik Blasch
Air Force Research Lab
Rome, NY, USA, 13441
erik.blasch.1@us.af.mil

***Abstract*—In this paper we present probabilistic graph fusion algorithms to support information fusion and reasoning over multi-source text media. Our methods resolve misinformation by combining knowledge similarity analysis and conflict identification with source characterization. For experimental purposes, we used the dataset of the articles about current military conflict in Eastern Ukraine. We show that automated knowledge fusion and conflict detection is feasible and high accuracy of detection can be obtained. However, to correctly classify mismatched knowledge fragments as misinformation versus additionally reported facts, the knowledge reliability and credibility must be assessed. Since the true knowledge must be reported by many reliable sources, we compute knowledge frequency and source reliability by incorporating knowledge provenance and analyzing historical consistency between the knowledge reported by the sources in our dataset.**

Keywords—multi-source fusion, graphical fusion, misinformation detection, open source exploitation, situation assessment, information wars, knowledge graph

I. Introduction

Detecting targets and predicting threats in high volume all-source intelligence is a central challenge in the Processing, Exploitation and Dissemination (PED) cycle. Traditionally, adversarial behaviors and activities of interest to intelligence analysts were detected using persistent Intelligence, Surveillance, and Reconnaissance (ISR) sensors. Today, the battlefields are shifting to “denied areas,” where the use of U.S. Military air and ground assets is limited. To succeed in new environments, the U.S. intelligence analysts increasingly rely on available open-source intelligence (OSINT), including local news, blogs, online investigations, and social media reports. These data often include multiple but slightly different (and sometimes conflicting) records of the same event or actors, and often contain missing, erroneous, and deliberately deceptive information. Compared to mainstream media, the social media provides particularly rich source of information as it represents real-time emotionally-charged and full of details coverage of events; however, these sources also contain vast amounts of biases and information manipulations.

The news delivery in social media is mostly tailored to the individual preferences of the consumers, which together with first-person style of reporting creates biased, partial and distorted information representation of the environment. Such properties of online and social media are exploited in social campaigns and media propaganda wars, particularly evident in the developing situations in Middle East and Eastern Europe. The ability of new OSINT sources to create “alternative reality” can cause state of denial and affect security situation in large areas, as particularly evident in the undeclared war waged by Russia against Ukraine, resulting in mass casualties, accompanied by wide-spread Russian government propaganda [9] [10] [23], and affected by the state of denial albeit moral support in Russian society. Even propaganda-based sources often contain information with unique details crucial to reconstructing the true situation on the ground, as evident by many online media investigations on the Russian-Ukrainian conflict that used multimedia and fact statements to geolocate events, infer controlled areas boundaries, and track military force movements. Thus it is essential for the impartial analysis to distinguish the real facts from the distortions and information manipulations.

To reconstruct the full state of the situation and track its evolution in time, the analysts must connect the events and entity mentions across time and multiple sources. Manual analysis of such data was always difficult in the past; with the increasing importance of social media as a primary source of information, manually analyzing these sources is infeasible due to its sheer volume. Analysts need tools for summarization and retrieval of information from OSINT sources, and these solutions must identify and resolve conflicting and deceptive information.

When combining multi-source data, it is critical to identify the claims from factual content. The reports in open source media often contain knowledge errors, manipulations and deceptions (e.g., due to propaganda) that are hard to identify due to ambiguity of linguistic expressions and uncertainty about identify and association of mentioned persons, organizations, locations, other physical objects, and events they are involved in. Information extraction tools only contribute to these challenges, as they produce high errors in extracting entities, relations and attributes from free-text data. For example, a state-of-the-art information extraction (IE) system working with 381,588 news documents from the Global Autonomous Language Exploitation (GALE) corpora produced more than 10 different incorrect country names to indicate where Osama Bin Laden was located.

978-1-4673-7557-3/15/$31.00 ©2015 IEEE

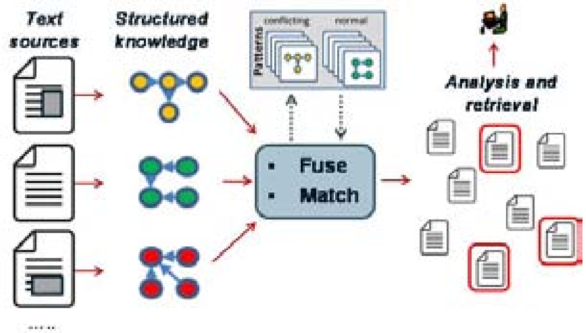

Fig. 1. Workflow of multi-source data analysis

In this paper we present a framework for a computational solution to assist OSINT analysts in detecting knowledge conflicts and identifying misinformation in text sources. Our solution is based to converting text into structured knowledge representation in the form of the *attributed graph*, and performing knowledge alignment, retrieval, conflict detect, and misinformation filtering using graph mining algorithms (Fig. 1). In the following, we start with a motivational example illustrating presence of misinformation across multiple text sources (Section II). We propose that multi-source data analysis and fusion from soft sources must incorporate misinformation detection in an integrated manner rather than bottom-up process. We summarize related research in information extraction, co-reference, and knowledge conflict detection in Section III, and describe the process of generating structured knowledge from unstructured text sources in Section IV. Section V illustrates how information fusion can be conducted by mapping the entities across different data sources, but reveals the challenges to "hard" fusion due to the presence of conflicting knowledge fragments or errors in the co-reference process. We then present an illustration of the joint information fusion and knowledge conflict detection. We define a formal model in Section VI that generalizes from presented examples to a computational algorithm that uses the graph mining, including subgraph pattern matching, learning, and fusion, as a core solution strategy. We conclude with discussion of next steps in Section VII.

II. Example of Misinformation in Multiple Text Sources

Some online reporting sources contain opinions rather than facts, and therefore are easy to identify and dismiss. Other online multimedia sources, especially propaganda controlled by government organizations or interest groups, use fake imagery or video as a proof of their situational analysis; such sources appear to draw increased emotional responses but are easier to detect thanks to extensive computer vision and imagery analysis tools available to analysts today. However, the most difficult misinformation to detect is often subtle: the sources producing such deceptions and information manipulations contain mostly accurate (although often irrelevant) information, and therefore cannot be easily detected and filtered.

In our research, we identified four classes of subtle information manipulations that are prevalent in online media and represent challenges to automated information analysis and fusion tools:

1) *changing the words when restating the information provided by other sources;*
2) *omitting important contextual information;*
3) *adding false or out-of-context statements; and*
4) *adding false associations or attributions.*

Chinese side deeply understands the challenges and hardships Russia faces in Ukrainian question and supports Russia's efforts to promote political solution to Ukrainian crisis on the basis of Minsk agreements.
Chinese side does not want to see military conflicts in Ukraine and suggests to solve Ukrainian crisis peacefully, using policy. Chinese side is against interference into internal policy of Ukraine by overthrowing Ukrainian political authorities.
Chinese side welcomes the participation of all Russian regions in Silk Road Economic Belt

Doc A: Original statements by Chinese authorities containing true information

BEIJING, November 21. /TASS/. China is against the declaration of independence by any ethnic groups through referendums, but this does not apply to Crimea, the acting director of the Chinese Foreign Ministry's European-Central Asian Affairs department, Gui Congyou, told Russian media on Friday.
"We should take a very careful and well-considered attitude to tackling nationalities' issues. We are against any nationality gaining independence through referendums. As far as Crimea is concerned, it has very special features. We know well the history of Crimea's affiliation," the diplomat said.
"On the whole, the nationalities' problems in some countries stem from double standard policies by certain states, which, proceeding from their own selfish interests, support one ethnic group and push it towards holding an independence referendum. This is a manifestation of double standards serving the interests of the United States. In a bid to achieve its aims, the US resorts to intervention in the internal affairs of other states by using force without UN Security Council authorization. China is firmly against this approach," Gui said.
In his opinion, "such actions trigger aggravations of inter-ethnic contradictions and result in armed conflicts."
On Russia's stance on crisis in Ukraine
China supports Russia's stance on settling the crisis in Ukraine, Gui Congyou stressed.
"China reacts with full understanding to the challenges and threats Russia has faced in connection with the Ukrainian issue and supports Moscow's approach to its settlement. We are not interested in an armed conflict on the Ukrainian territory and wish to see the issue settled by political means. We are against external intervention in Ukraine's internal affairs through government coups."
"As for the causes of the Ukrainian crisis, in a telephone conversation with Russian President Vladimir Putin the Chinese leader, Xi Jingping said that 'there is no smoke without fire'," Gui added.
"Personally, I believe that some states and blocs stick to the Cold War mentality and are involved in geopolitical games with the aim to infringe on the geopolitical interests of Russia. China is strongly against anyone harming the security of other countries for the purpose of protecting one's 'absolute security'," he noted.

Doc B: News stories by ITAR-TASS containing information misrepresentation

Fig. 2. Example of information manipulation

Fig. 2 shows an example that contains all four of the abovementioned manipulations. The situation involved a set of original statements made by Chinese officials (Document A), as translated from Chinese foreign ministry website, and a report by Russian government ITAR-TASS news agency (Document B) which misrepresented the statements by Chinese authorities. The examples of information manipulation include:

- *changing the words*: when quoting Chinese authorities, ITAR-TASS made a subtle change of the word from "hardship" to "threat". The new word carries more emotional response and skews the meaning of the statement: this change is intended to convince the readers that Chinese authorities agree that Ukraine is a threat to Russia, and that they agree with Russia's approach to solving Ukrainian crises. This manipulation could be detected by finding the alignment between the knowledge fragments in original source and Russian quotes and a corresponding mismatch between the words used to describe the same information object.

- *omitting important contextual information*: original Chinese's statement "Chinese side … supports Russia's efforts to promote political solution to Ukrainian crisis on the basis of Minsk agreements." was paraphrased by the ITAR-TASS as "China … supports Moscow's approach to its settlement.", and the reference to Minsk agreements was omitted. This change seems to suggest that China supports the Russia's actions in Ukraine, even possibly including military actions, which is not true. This manipulation can be detected by finding the absence of the information fragment directly attached to the reported information fragment and aligns with original source.

- *adding false or out-of-context statements*: ITAR-TASS added several statements in its article that may have been taken out of different context but did not appear in original statements by Chinese authorities, including "*As far as Crimea is concerned, it has very special features. We know well the history of Crimea's affiliation*.", and "*We are against external intervention in Ukraine's internal affairs through government coups*". The first statement subtly makes an appearance that Chinese officials, while being against separatism, consider Crimea to be a different case and agree that it has Russian affiliation. The second statement is intended to give readers a feeling that China agrees with Russia's position that government in Ukraine was changed via a coup. Both statements are subtle and influence the readers on the emotional level, but are ambiguous enough to not stand in libel court.

- *adding false associations or attribution*: ITAR-TASS published a reference to China's official opinion regarding United States policies of supporting independence of ethnic groups. The article also cites a statement that "*such actions trigger aggravations of inter-ethnic contradictions and result in armed conflicts*". This is subtly implying that China agrees that armed conflict in Ukraine is caused by US's interference. This association, which is an official position of Russian Federation, is incorrectly attributed to Chinese officials. Detection of false associations can be done by aligning the entities in different reports and identifying the relations or paths in knowledge graph between entities that are present in one report but absent in another source.

In this paper, we describe computational model to analyze knowledge reported in multiple sources, and discover four types of information conflicts described above. We note that the implicit association links, like the example above, are hardest to detect because alignment of knowledge from two sources with each other is not enough to discover these links. An algorithm that can detect such associations must use supplemental information, such as the prevalent opinions of the readers of the news media. While this information can be inferred from analyzing the history of the news media reports and the comments and responses of its audience, it represents our future research and is beyond the scope of this paper.

III. Related Research

Developing automated knowledge analysis and fusion algorithms for unstructured text requires it to first be converted into a structured representation. This is usually performed by IE tools that generate entities (people, groups, places, events), relations among entities, and series of events associated with an entity or groups of entities. The information extraction problem has been artificially broken down into several components such as entity mention boundary identification, entity type classification, relation extraction, and event detection and categorization. Although adopting such a pipelined approach would make a system comparatively easy to assemble, it has many limitations. The errors in upstream component, such as mislabeling of an entity or relation type, are often compounded and propagated to the downstream classifiers without any feedback [22]. Recent research in joint entity, relation, and event extraction resolves these limitations by exploiting global features that explicitly capture inter-dependencies among multiple event triggers and arguments [21].

Even after IE tools produce structured knowledge representations, fusion of multiple documents cannot rely on known object identities, and the features that allow low-ambiguity comparison between multiple mentions of the same object or event, such as time and geolocation features, are rarely present. As the result, the text fusion requires solving cross-document entity co-reference problem, which must be executed jointly to increase the co-reference accuracy [25] [26]. The knowledge of which events co-refer can help disambiguate entities, and vice versa, hence joint entity and event co-reference will achieve higher association accuracy [20].

Knowledge conflict detection and resolution have been recently explored in online deception detection, knowledge base population, and textual entailment and contradiction detection domains. One of the active research areas involves analyzing the linguistic and stylistic properties of textual records [1] [2] [15] [31]. Authorship attribution and stylistic cues have been extensively used in online fraud detection, such as detecting sockpuppets in Wikipedia [28], finding deceptive online profiles [30], discovering reputation fraud, social and opinion spam [27]. State-of-the-art methods usually combine rule-based feature specification with statistical classifiers, and make decisions at a high level (e.g., a user or an article). Instead, our objective is to identify false knowledge hidden within valid knowledge, which is often the case due to obfuscation behaviors of the sources. This requires detailed analysis of the knowledge fragments, their relations, associations and similarities to other available knowledge.

In treating the conflict patterns as semantic graphs of the entities and relations, our work is closely related to knowledge base population, joint co-reference, and global reasoning [7] [13] [14] [16] [20] [26]. However, these models usually define small patterns [20] or local rules, which cannot describe the conflicts across multiple records. Our approach to multi-source knowledge conflict detection is based on constructing global cross-record conflict patterns learned in the data corpus using the local conflict rules [18]. Our patterns and instance-based conflict exemplars maintain cross-document association uncertainty as probabilistic distributions. This allows us to avoid hard co-reference decisions that may introduce errors [14]. We then fuse multiple sources by constructing "soft" co-references, perform retrieval as subgraph detection in this aggregated graph, and generate knowledge frequency estimates using source identification information.

IV. GENERATING STRUCTURED KNOWLEDGE FROM UNSTRUCTURED TEXT

The model for multi-source information fusion presented in this paper converts text articles into attributed knowledge graphs from the products of IE tool. In our research, we used several IE tools, including SERIF [3], Stanford dependency parser [6], and Semantex [29])[1]. We use the information contained in IE outputs to construct nodes, relations, and their attributes. Resulting graph contains both semantic and syntactic information embedded in the original text, and is somewhat similar to the structure of proposition graphs [24].

Fig. 3. Examples of tokenization output from Semantex

To motivate mathematical definitions, we illustrate the process of knowledge graph creation using the outputs of Semantex tool. Semantex generates token hierarchy, named entities, profiles, relations, and events:

- **Token hierarchy** (Fig. 3) includes all of the word segments extracted by parsing the text in the document. Tokens are groups of words and represent candidates to be converted to nodes in the knowledge graph. As we want the graphs to be compact and comparable, we need to make decisions at what level of token hierarchy to generate the graph nodes. Processing document A from Fig. 2 with 3 sentences, Semantex generated 42 top-level tokens (>100 "atom" tokens). For some of the atom (leaf) tokens Semantex generates syntactic features (such as part-of-speed tag), semantic features (e.g., WordNet hyperhyms, ontology references), and word descriptors (Fig. 4c).
- **Named entities** are tokens representing mentions of known people, locations, organizations, and geopolitical entities in the text. For the sentence in Fig. 3, Semantex identified three named entities: Russia, Chinese side, and Minsk
- **Profile** list is generated as co-referenced entity mentions. Profile list helps make decision on the token-to-node conversion and token hierarchy cutoff: each profile becomes a node in the knowledge graph. For Document A, Semantex extracted seven entity profiles.
- **Event list** contains activities detected by the tool. Events are usually verb phrases connecting the non-verb (usually noun) tokens. Each event becomes a node in the knowledge graph. For the sentence above, there were two events extracted by Semantex (Fig. 4 shows a structure of event "*promote*").
- **Relations**: there are a number of syntactic relations that Semantex generates that are useful for our knowledge graph construction, including "possessive", "has description", "purpose", "accompanying contextual", "location", etc. (Fig. 5). We also use traditional verb-object/subject to supplement event detections which sometimes miss the subject-object relations or contain ambiguous linking (e.g., entity is labeled as "*participating*" in the event detection, while it actually represents the *location* of the event).

The brute-force method to extract the knowledge graph from Semantex output is to generate all tokens as nodes, add the links corresponding to the token hierarchy, events, and relation objects, and then attempt to aggregate this to a more compact representation. Fig. 6a shows a full network generated using the brute-force method, and a subgraph corresponding to the sentence above. As we can see, the full graph is too dense and confusing, with multiple nodes corresponding to the same entity and others not carrying much information. Consequently, we developed a process for aggregating tokens: first, all tokens associated with the same profile are combined into a single node; second, all non-verb tokens on the same path in token hierarchy are aggregated into a single node. The resulting knowledge graph is compact and unambiguous representation of the knowledge contained in the sentence (Fig. 6b).

While we are able to generate compact knowledge graphs, the quality of the graph still depends on the quality of the IE outputs. For the example shown in Fig. 6, we notice that Semantex missed the "verb-subject" relation between "Chinese side" entity and "support" event. This relation is correctly extracted by Stanford typed relation parser, and it is our experience that combining the outputs from multiple IE tools will increase the quality of constructed knowledge graphs.

[1] http://www.prweb.com/releases/2011/8/prweb8708189.htm

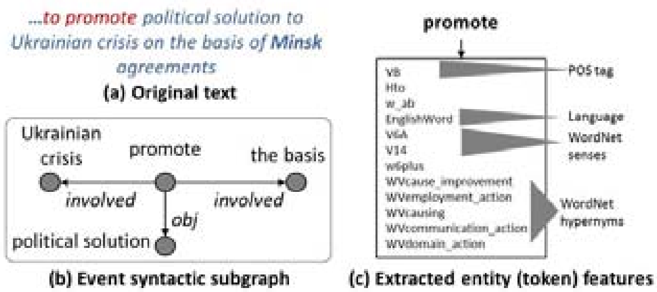

Fig. 4. Examples of event "to promote" extracted by Semantex

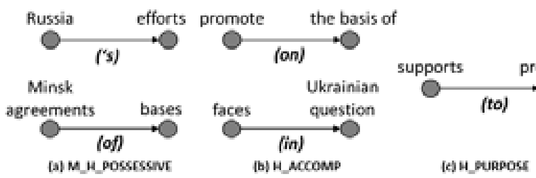

Fig. 5. Examples of relations extracted by Semantex

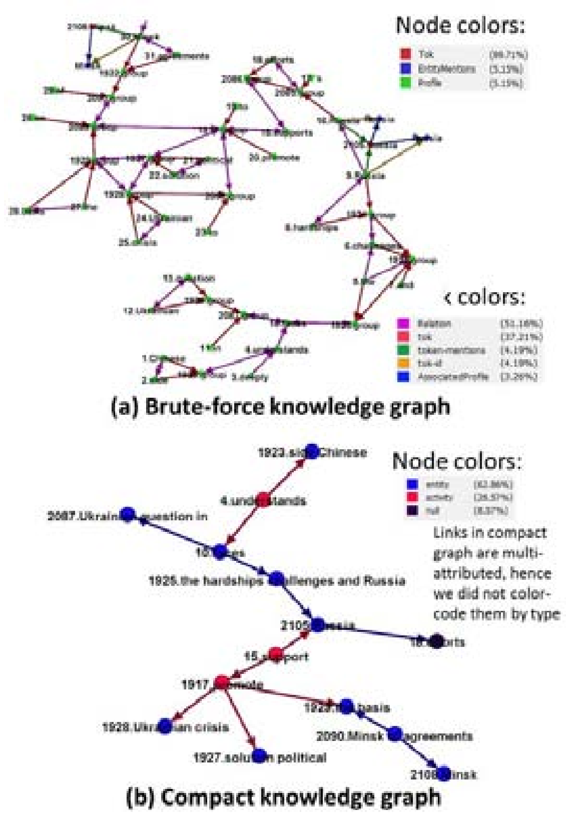

Fig. 6. Knowledge graph extracted for the sentence "*Chinese side deeply understands the challenges and hardships Russia faces in Ukrainian question and supports Russia's efforts to promote political solution to Ukrainian crisis on the basis of Minsk agreements*"

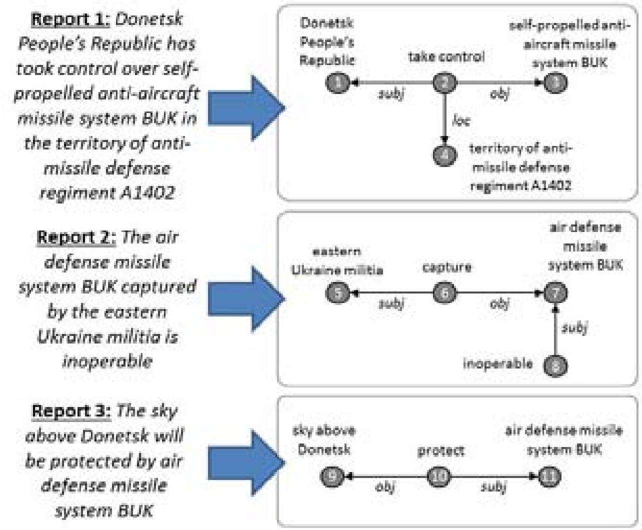

Fig. 7. Examples of knowledge extraction

V. Knowledge Alignment and Conflict Analysis using Node Mapping

Fusing or comparing knowledge from multiple text reports is challenging because only few of the tokens, and correspondingly the nodes in the knowledge graphs, have identity information. The identity is only available in the nodes corresponding to the profiles, and all other event and token nodes must be mapped to make knowledge comparison decisions. To illustrate this problem, consider the knowledge graphs generated from three text reports (Fig. 7). In the first report, the press release from Donetsk People's Republic (DPR), which was shared by multiple Russian news agencies, claimed that DPR took control over the surface-to-air missile system called "BUK" when their militia stormed the barracks of anti-missile defense regiment A1402. In the second report, the Ukrainian security services declared that the system captured by militia is inoperable. In the third report pro-rebel social media sources claimed that the Eastern Ukrainian militia will establish a protection of its air space using BUK system. These were just three of the many stories surrounding the downing of the Malaysian civilian airliner MH17 in summer of 2014 over Eastern Ukraine.

A. Knowledge alignment and fusion

Consider the first two reports: assuming they both are true, it may be easy for a human reader to conclude that the BUK system captured by Eastern Ukraine militia is inoperable. However, for automated solution to make the same conclusion, which can be done by creating a fused knowledge from the partial information available in each of the sources, the model has to establish a correspondence, or *mapping*, between the entities mentioned in both sources. An example of such mapping is shown in Fig. 8a. First, we map "*Donetsk People's Republic*" (DPR) entity from report 1 to "*eastern Ukraine militia*" entity in report 2. Due to the recency of the conflict in Eastern Ukraine, this mapping is ambiguous, as there were in fact separate militia groups operating in the region at the time. Next, due to semantic equivalence, we can map event "*take control*" from report 1 to event "*capture*" in report 2. Note that this mapping is also ambiguous: while the report 1 indicates a location where the event occurred, no such information is provided in report 2. Therefore, it is quite possible that report 2 refers to a different instance of the even of the same type as the event in report 2, and further analysis may be needed to establish that this mapping is valid. Finally, we map entity "*self-propelled anti-aircraft missile system BUK*" from report 1 to entity "*air defense missile system BUK*" in report 2. While this mapping appears valid, there are three challenges that automated knowledge fusion model has to face:

- Matching the group of descriptor words "*self-propelled anti-aircraft*" with "*air defense*" involves non-trivial linguistic inference: (i) the first entity description discusses the mobility property of the system ("*self-propelled*") while the second does not contain this information, and (ii) the first contains "anti-aircraft" as a reference to the purpose of the system, while the second contains reference to "*air defense*", not to be confused with "defense against bad air" that may be presumed using syntactic information alone;

- BUK system actually contains 4 subcomponents[2]: acquisition radar, command component to discern "friendly" military aircraft from foes and pass radar targeting information to the missile launchers, missile launcher component, and logistics component. Neither of the reports mention which of the components (or all four) are meant; it is feasible that report 1 intended to inform about launcher component while report 2 mentions command component.

- The investigations of Ukrainian conflict revealed that Eastern Militia are in possession of multiple BUK systems, and it cannot be presumed from these reports whether they are about the same system, since the first report informs about the location that militia took control of the system, while the second does not. The human reader derives this contextual relation implicitly, but the automated system must do this explicitly; moreover, the implicit nature of many comprehension tasks is a rich space for various information manipulations.

Entity mapping described above is called *cross-document co-reference* [16]. Assuming that entity mapping is accurate, we can generate a fused knowledge that "*DPR captured BUK system at the territory of regiment A1402 that is inoperable*" represented structurally as a fused graph in Fig. 8b.

B. Knowledge conflicts

If we apply the same fusion process to all three reports, we may obtain the statement that "*DPR captured BUK system at the territory of regiment A1402 that is inoperable and will protect the DPR's air space*", which the corresponding graph depicted in Fig. 8c. A human observer immediately identifies this knowledge as inherently flawed: an air-defense system that is inoperable cannot be used for protection of the airspace.

[2] http://en.wikipedia.org/wiki/Buk_missile_system

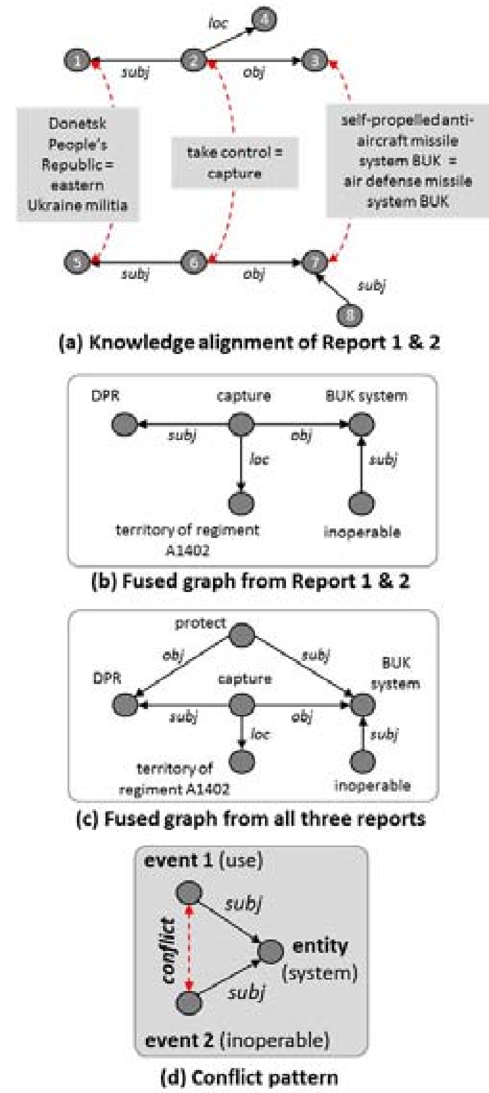

Fig. 8. Examples of knowledge alignment and a corresponding fused graph

There could be several possibilities why this knowledge is in conflict:

- <u>One of the reports that was used during fusion process contained false information:</u> The investigations around Eastern Ukrainian conflict revealed that in fact all reports were accurate: the militia did storm the base, did get in the possession of the BUK system, but the system was disabled by retreating Ukrainian military units evident from the photos shown to the reporters. Further, multiple Ukrainian military transport jets have been downed by militia, and the situation culminated with the downing of the Malaysian civilian airliner MH17 at the altitude above 10,000 km.

- <u>Some event occurred during the time between the reports, rendering information in one of them obsolete:</u> It was initially hypothesized that the DPR could repair the BUK system. However, Google Earth imagery confirmed that the captured BUK component remained at the same location of A1402 regiment throughout the unfolding events of summer, indicating that the system was never operated.

- <u>The fusion process established incorrect mapping between the entities in different reports:</u> From the investigations it appears that the BUK system(s) used to attack Ukrainian jets and civilian airplane were not the one described in Report 1. Numerous online investigations analyzing available imagery, audio, video, and witness statements (e.g., see Bellingcat's reports[3]) concluded that the BUK system used to down MH17 was fired from rebel-controlled territory, and was mostly likely delivered from Russia. In fact, the Reports 1 and 2 were intended to support a storyline that militia captured BUK system (rather than got them from Russia).

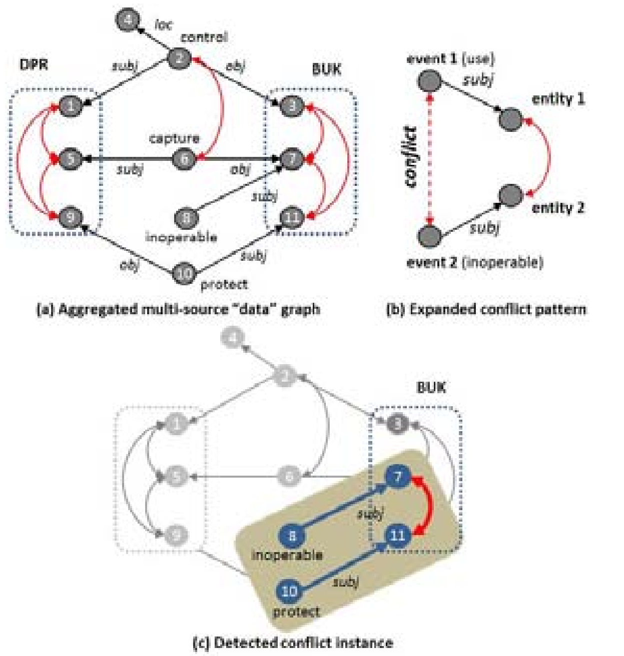

Fig. 9. Aggregated data graph, expanded conflict pattern, and detected knowledge conflict

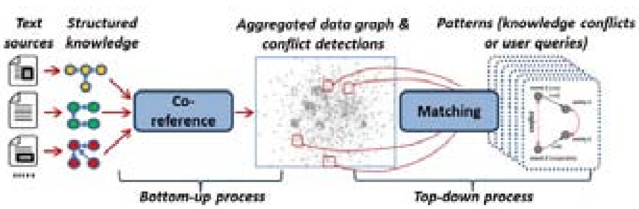

Fig. 10. Conflict detection using knowledge matching

C. Integrating fusion and conflict detection

How can we automate detection of such knowledge conflicts? On further inspection, we see that the conflict described in the above example can be generalized as a simple triangular pattern of an entity being a subject (agent) of two events that are in conflict with each other. In our case the entity is BUK system with its multiple mentions across the reports, and the conflicting events are the *inoperability* of the BUK system and its *use to protect* (shoot down airplanes) the airspace of the area under control by militia in Eastern Ukraine. We have shown previously that the patterns like this, as well as more complex knowledge conflict patterns, can be constructed by using a combination of local conflict rules, global contextual and co-reference relations, and graph learning [18]. These patterns thus combine local and global features and encode the relative source relationships in their structure.

To detect knowledge conflicts using these patterns, we introduce a concept of *aggregated data graph*. It is constructed by combining all knowledge graphs from multi-source report graphs, and generating "soft" entity co-reference links (red links in Fig. 9a). Next, we expand the conflict patterns to encode possible cross-document co-reference by generating copies of the nodes with multiple relations to other nodes (Fig. 9b). Finally, we match these expanded patterns to the aggregated data graph to retrieve similar subgraphs (Fig. 10). An example of detected cross-document knowledge conflict is depicted in Fig. 9c. Similar process is followed if we need to perform queries against multi-source data: in this case, the patterns represent queries.

VI. KNOWLEDGE ANALYSIS MODEL

We are finally ready to introduce a formal computational model and algorithms that can be used to detect knowledge conflicts and generate fused knowledge from multiple text sources.

A. Structured knowledge representation using attributed graphs

Formally, a *knowledge graph* is defined as attributed graph $G = (V, E, A)$, where $V = \{1, \dots, M\}$ is a set of vertices (representing entities and events), $E = \{(k, m); k, m \in V\}$ is a set of edges (representing semantic and syntactic relations), and $A = \{a_{km}\}$ are attributes such as local semantic, syntactic, and linguistic features [22] describing the entities and their relations, i.e. a_{kk} are attributes of entity/event k, and a_{km} are attributes of the relation (k, m) between entities k and m (Fig. 11). The patterns, which we call *model graphs*, can be denoted as $\mathbf{G^M} = (V^\mathbf{M}, E^\mathbf{M}, A^\mathbf{M})$, while aggregated knowledge, called *data graph*, is denoted as $\mathbf{G^D} = (V^\mathbf{D}, E^\mathbf{D}, A^\mathbf{D})$.

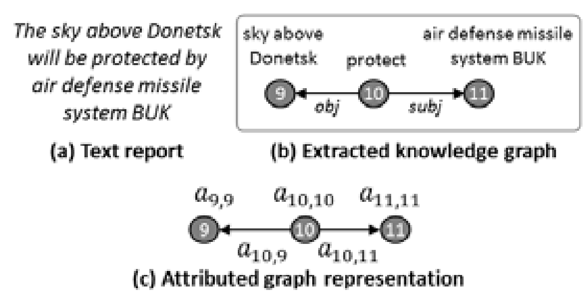

Fig. 11. Text is converted into attributed graphs

[3] https://www.bellingcat.com/tag/mh17/

B. Graph node-to-node mapping

As discussed previously, the knowledge alignment, fusion, and conflict assessment require performing mapping of the nodes in knowledge graphs of corresponding sources. We define the mapping as a 0-1 assignment matrix $X = \{x_{ki}\}, x_{ki} \in \{0,1\}$ of the nodes $V^{\mathbf{M}}$ in model graph to the nodes $V^{\mathbf{D}}$ in the data graph (Fig. 12a), with more than one model node possibly mapped to the same data node (this is equivalent to assuming that multiple entity mentions encoded in different nodes in the model graph are resolved in the same data entity or event instance).

We define a score of the match between the model and data graph $\varepsilon(X|M,D) = e^{-\frac{Q(X)}{\eta(\mathbf{G}^{\mathbf{M}})}}$, where $Q(X)$ is a quadratic function of the mismatch between the model graph and mapping-induced subgraph in the data, and the normalization coefficient $\eta(\mathbf{G}^{\mathbf{M}})$ corresponds to the norm of the model (normalization is needed to compute relative and meaningful detection scores, as well as disambiguate the data nodes that seem to match to different queries or knowledge conflict patterns):

- Mismatch: $Q(X) = \underbrace{\sum_{ki} x_{ki} w_{ki}}_{node\ mismatch} + \underbrace{\sum_{kmij} x_{ki} x_{mj} w_{kmij}}_{link\ mismatch}$
- Model norm: $\eta(\mathbf{G}^{\mathbf{M}}) = \underbrace{\sum_k w_{k,null}}_{node\ norm} + \underbrace{\sum_{km} w_{km,null}}_{link\ norm}$

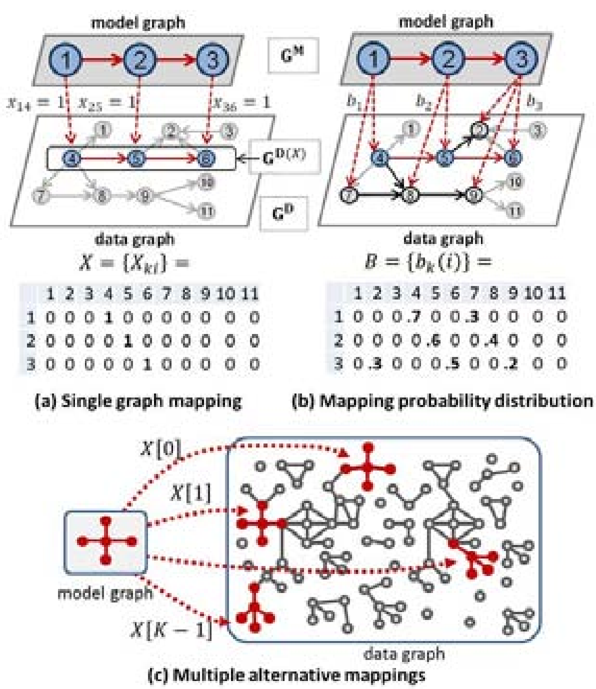

Fig. 12. Graph mapping and its variables

In the above, parameters w_{ki} represent weight of *mismatch* between node attributes $a_{kk}^{\mathbf{M}}$ and $a_{ii}^{\mathbf{D}}$, and parameters $w_{km,ij}$ represent mismatch between link attributes $a_{km}^{\mathbf{M}}$ and $a_{ii}^{\mathbf{M}}$, – of models and data graphs respectively. Node and link mismatches have probabilistic interpretation: they define the negative log-likelihood density function of observing the attributes $a_{i,j}^{\boldsymbol{D}}$ when the true attributes are $a_{k,m}^{\boldsymbol{M}}$. In practice, these attribute mismatch coefficients can be computed using a range of functions, distributions and local models. We most often use the L2 norm, assuming the Gaussian noise in the attribute values, and the semantic distance between word descriptors.

The matching between user-defined query or knowledge conflict pattern and aggregated data graph is possible because we can compute the mismatches between nodes and links in the attributed graphs. When dealing with text data, such mismatches need to be adapted to incorporate the semantic information about the descriptors of corresponding entity and relation pairs. In our research, we used three metrics to compare entity and relation mentions: (i) class distance, (ii) synset distance, and (iii) edit distance. Class and synset distances use information from WordNet database[4]. Hop distance measures the semantic distance in terms of class similarity. For example, the word pair *dog* and *cat* are both animals and would be closer than the pair *dog* and *house*. Synset distance measures the synonym distance between two words. All synonyms of a word in WordNet are linked to the same synset cluster. The length of a path in synset network between synset clusters of corresponding words assesses the similarity of the words in terms of their synonymy. Finally, edit distances allows to account for word typos and alternative spellings.

C. Detection and search via mapping estimation

Consequently, the conflict detection or multi-source queries can be performed by finding the subgraphs of the data graph and corresponding node assignments $X = \{x_{ki}\}$ for which the function $\varepsilon(X|M,D)$ is minimized. For a pattern graph $\mathbf{G}^{\mathbf{M}}$, the we generate a query output or a conflict detection output as a collection of tuples $\langle X, \mathbf{G}^{\mathbf{D}(X)} \rangle$, where X is the mapping and $\mathbf{G}^{\mathbf{D}(X)}$ is the corresponding data subgraphs for which $\varepsilon(X|M,D)$ is above threshold. The data subgraph $\mathbf{G}^{\mathbf{D}(X)} = \left(V^{\mathbf{D}(X)}, E^{\mathbf{D}(X)}, A^{\mathbf{D}(X)}\right)$ is ***induced*** by the mapping X, i.e. $V^{\mathbf{D}(X)} = \{i \in V^{\mathbf{D}}: \sum_{k \in V^{\mathbf{M}}} x_{ki} = 1\}$, and $E^{\mathbf{D}(X)} = \{(i,j) \in E^{\mathbf{D}}: \sum_{k \in V^{\mathbf{M}}} x_{ki} = \sum_{k \in V^{\mathbf{M}}} x_{kj} = 1\}$.

Traditionally, attributed graph matching problem has been solved by finding the mapping $X^* = \{x_{ki}^*\}, x_{ki}^* \in \{0,1\}$ to minimize the mismatch between attributes of mapped nodes of graphs $\mathbf{G}^{\mathbf{M}}$ and $\mathbf{G}^{\mathbf{D}}$: $X^* = \arg\min \varepsilon(X|M,D)$. We have developed graph matching algorithm that, instead of finding a single-best matching X^* (Fig. 11a), finds an estimate of marginal posterior probability distributions of the node-to-node matching (Fig. 11b): $b_k(i) = \Pr(x_{ki} = 1|M,D)$. These probabilities optimize an energy function that is composed of the joint probability of graph mapping variables and the entropy of the mapping alternatives. This allows us to find multiple graph matches that are close to the optimal solution (Fig. 11c), representing multiple query responses and/or

[4] http://wordnet.princeton.edu/

conflict detections. Our graph matching algorithm, based on smoothed loopy belief propagation [17], iteratively updates the estimates of mapping probabilities by passing belief messages in the factor graph constructed using topological information in the model and data graphs [19].

D. Knowledge frequency estimation

Knowledge conflict detection is just the first step in detecting misinformation. Each conflict detection output $\langle X, \mathbf{G}^{\mathbf{D}(X)} \rangle$ consists of knowledge fragments (entities, events, and relations) from different records/sources that occur in the induced graph(s) $\mathbf{G}^{\mathbf{D}(X)}$. Decisions can then be made which fragment represents false knowledge by analyzing three hypotheses:

- *Hypothesis 1: The true knowledge must be provided by many sources*
- *Hypothesis 2: The true knowledge is reported by trust-worthy sources*
- *Hypothesis 3: The true knowledge must be provided by independent sources*

In our research, we currently work on estimating source reliability and credibility (hypothesis 2) and independence (hypothesis 3), hence this work is outside of the scope for this paper. Hypothesis 1 can be evaluated by estimating a frequency of the knowledge fragment $\mathbf{G}^{\mathbf{D}(X,s)} \subset \mathbf{G}^{\mathbf{D}(X)}$ from the source s. This can be achieved by performing matching of the subgraph $\mathbf{G}^{\mathbf{D}(X,s)}$ against all other reports' graphs $\mathbf{G}^{\mathbf{D}[u]}$ for all sources u, and then counting the sources for which the objective function $\varepsilon(s,u) = \arg\max_Y \varepsilon(Y|\mathbf{D}(X,s), \mathbf{D}[u])$ is above the threshold. This frequency is equivalent to an estimate of *plausibility* of knowledge content based on its consistency with the knowledge reported by other sources. The knowledge fragments with lowest plausibility scores in the conflict detection subgraphs $\mathbf{G}^{\mathbf{D}(X)}$ can be declared as false and either removed from the dataset, or provided to the users for further investigation.

VII. Conclusions

In this paper we illustrated the challenges working with OSINT data using real-world examples of news and social media reports from Russian-Ukrainian conflict. The multiple sources available for exploration and fusion often contain ambiguous and manipulated information, and hence multi-source knowledge fusion process must detect, classify, and resolve misinformation and deceptions. Our experiments with this dataset showed feasibility of increasing the quality of multi-source fusion and accuracy of cross-document misinformation detection. We did notice a significant dependency of the quality of conflict detection and classification on the accuracy of relation and event extraction. Since multiple different IE tools exist on the market and in open source community, it is our assessment that fusing their outputs, rather than developing a new IE solution, would result in generation of more complete and meaningful semantic knowledge graphs.

Knowledge conflict detection may require reasoning about explicitly stated and implied knowledge. This can be addressed by textual entailment [4], which discovers if one text record is highly likely true given that another record is true. Common approaches to textual entailment often rely on comparing textual features, such as tokens, subsequences, numbers, dates, and named entities between a pair of documents [8] [12]. However, textual entailment can only analyze two records at a time, and feature-based approaches used to detect entailments do not work as well for complex knowledge contexts where semantic relations become critical. The use of texts' structure for contradiction detection has been previously explored through the alignment of dependency parse trees and consequent contradiction scoring using a series of pairwise token comparison rules, such as antonymy, negation, numeric, and date mismatch [6].

Another important topic for research is estimating the independence of different sources. We can use the reposting behaviors and comment attitudes in news and social media to estimate the topic independence. Then establishing the provenance of the knowledge fragments reported by the source over time [11] will allow assessing when the source generates original content, how different this content is from other sources, and track the similarity in both behavior and content to estimate the source's pedigree. We plan to address these challenges in our current and future research efforts.

References

[1] Afroz, S., Brennan, M., & Greenstadt, R. (2012, May). Detecting hoaxes, frauds, and deception in writing style online. In Security and Privacy (SP), 2012 IEEE Symposium on (pp. 461-475).

[2] Arciuli, J., Villar, G., & Mallard, D. (2009). Lies, lies and more lies. In Proceedings of the 31st Annual Conference of the Cognitive Science Society (CogSci 2009) (pp. 2329-2334).

[3] Boschee, E., Weischedel, R., & Zamanian, A. (2005, May). Automatic information extraction. In Proceedings of the 2005 International Conference on Intelligence Analysis, McLean, VA (pp. 2-4).

[4] Dagan, I., Dolan, B., Magnini, B., Roth, D., Dagan, I., Dolan, B., & Pantel, P. (2010). Recognizing textual entailment: Rational, evaluation and approaches–erratum. Natural Language Engineering, 16(1), 105.

[5] De Marneffe, M. C., MacCartney, B., & Manning, C. D. (2006, May). Generating typed dependency parses from phrase structure parses. In Proceedings of LREC (Vol. 6, No. 2006, pp. 449-454).

[6] De Marneffe, M. C., Rafferty, A. N., & Manning, C. D. (2008). Finding Contradictions in Text. In ACL (Vol. 8, pp. 1039-1047).

[7] Dredze, M., McNamee, P., Rao, D., Gerber, A., & Finin, T. (2010, August). Entity disambiguation for knowledge base population. In Proceedings of the 23rd International Conference on Computational Linguistics (pp. 277-285). Association for Computational Linguistics.

[8] Ghuge, S., and Bhattacharya, A. (2013) Survey in Textual Entailment. Center for Indian Language Technology, retrieved on April, 2014, from http://www.cfilt.iitb.ac.in/resources/surveys/TE-Literature%20Survey-2013-Swapnil_Ghuge.pdf

[9] Goolsby, R. and N. Agarwal (2014), "Information Tactics and Maneuvers in the New Information Environment", In the NATO Systems Analytics Symposium, November

[10] Gregory, P.R. (December 2014) "Putin's New Weapon In The Ukraine Propaganda War: Internet Trolls", Forbes, Dec 09, 2014. Retrieved on Feb 27, 2015 from http://www.forbes.com/sites/paulroderickgregory/2014/12/09/putins-new-weapon-in-the-ukraine-propaganda-war-internet-trolls/

[11] Hartig, O. (2009, April). Provenance Information in the Web of Data. In LDOW.

[12] Hickl, A. (2008, August). Using discourse commitments to recognize textual entailment. In Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1 (pp. 337-344). Association for Computational Linguistics.

[13] Hong, Y., Zhang, J., Ma, B., Yao, J., Zhou, G., & Zhu, Q. (2011, June). Using cross-entity inference to improve event extraction. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1 (pp. 1127-1136). Association for Computational Linguistics.

[14] Ji, H., Deng, H., & Han, J. (2012). Uncertainty Reduction for Knowledge Discovery and Information Extraction on the World Wide Web. Proceedings of the IEEE, 100(9), 2658-2674.

[15] Juola, P. (2012, April). Detecting stylistic deception. In Proceedings of the Workshop on Computational Approaches to Deception Detection (pp. 91-96). Association for Computational Linguistics.

[16] Lee, H., Recasens, M., Chang, A., Surdeanu, M., & Jurafsky, D. (2012, July). Joint entity and event coreference resolution across documents. In Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (pp. 489-500). Association for Computational Linguistics.

[17] Levchuk, G., J. Roberts, and J. Freeman (2012) "Learning and Detecting Patterns in Multi-Attributed Network Data", AAAI Fall Symposium on Social Networks and Social Contagion.

[18] Levchuk, G., Jackobsen, M., & Riordan, B. (2014, June). Detecting misinformation and knowledge conflicts in relational data. In SPIE Defense+ Security (pp. 90910P-90910P). International Society for Optics and Photonics.

[19] Levchuk, G., Ortiz, A., & Yan, X. (2014, May). Analysis of large-scale distributed knowledge sources via autonomous cooperative graph mining. In SPIE Sensing Technology+ Applications (pp. 91190K-91190K). International Society for Optics and Photonics

[20] Li, Q., Anzaroot, S., Lin, W. P., Li, X., & Ji, H. (2011, October). Joint inference for cross-document information extraction. In Proceedings of the 20th ACM international conference on Information and knowledge management (pp. 2225-2228). ACM.

[21] Li, Q., & Ji, H. (2014). Incremental joint extraction of entity mentions and relations. In Proceedings of the Association for Computational Linguistics.

[22] Li, Q., Ji, H., & Huang, L. (2013, August). Joint Event Extraction via Structured Prediction with Global Features. In ACL (1) (pp. 73-82).

[23] Ormiston, S. (2015) "Russian government propaganda stokes anti-West sentiment", CBS News, Feb 04, 2015 retrieved from http://www.cbc.ca/news/world/russian-government-propaganda-stokes-anti-west-sentiment-1.2944274

[24] Prentice, M., & Shapiro, S. C. (2011, July). Using propositional graphs for soft information fusion. In Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on (pp. 1-7). IEEE.

[25] Singh, S., Subramanya, A., Pereira, F., & McCallum, A. (2011, June). Large-scale cross-document coreference using distributed inference and hierarchical models. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1 (pp. 793-803). Association for Computational Linguistics.

[26] Singh, S., Riedel, S., Martin, B., Zheng, J., & McCallum, A. (2013, October). Joint inference of entities, relations, and coreference. In Proceedings of the 2013 workshop on Automated knowledge base construction (pp. 1-6). ACM.

[27] Solorio, T., Hasan, R., & Mizan, M. (2013, June). A case study of sockpuppet detection in wikipedia. In Workshop on Language Analysis in Social Media (LASM) at NAACL HLT (pp. 59-68).

[28] Song, L., Zhang, W., Liao, S. S., & Kwok, R. C. W. (2012). A Critical Analysis Of The State-Of-The-Art On Automated Detection Of Deceptive Behavior In Social Media. In PACIS (p. 168).

[29] Srihari, R. K., Li, W., Niu, C., & Cornell, T. (2003, May). Infoxtract: A customizable intermediate level information extraction engine. In Proceedings of the HLT-NAACL 2003 workshop on Software engineering and architecture of language technology systems-Volume 8 (pp. 51-58). Association for Computational Linguistics.

[30] Toma, C. L., & Hancock, J. T. (2010). Looks and lies: The role of physical attractiveness in online dating self-presentation and deception. Communication Research, 37(3), 335-351.

[31] Zhou, L., Twitchell, D. P., Qin, T., Burgoon, J. K., & Nunamaker Jr, J. F. (2003, January). An exploratory study into deception detection in text-based computer-mediated communication. Proceedings of the 36th Annual Hawaii International Conference on System Sciences (p. 10)