A Semantic Approach for Automating Knowledge in Policies of Cyber Insurance Services

Ketki Joshi
Department of Information Systems
University of Maryland,
Baltimore County
Baltimore, MD, USA, 21250
Email: ketkij1@umbc.edu

Karuna Pande Joshi
Department of Information Systems
University of Maryland,
Baltimore County
Baltimore, MD, USA, 21250
Email: karuna.joshi@umbc.edu

Sudip Mittal
CSEE Department
University of Maryland,
Baltimore County
Baltimore, MD, USA, 21250
Email: smittal1@umbc.edu

***Abstract*—With the rapid adoption of web services, the need to protect against various threats has become imperative for organizations operating in cyberspace. Organizations are increasingly opting to get financial cover in the event of losses due to a security incident. This helps them safeguard against the threat posed to third-party services that the organization uses. It is in the organization's interest to understand the insurance requirements and procure all necessary direct and liability coverages. This helps transfer some risks to the insurance providers. However, cyber insurance policies often list details about coverages and exclusions using legalese that can be difficult to comprehend. Currently, it takes a significant manual effort to parse and extract knowledgeable rules from these lengthy and complicated policy documents. We have developed a semantically rich machine processable framework to automatically analyze cyber insurance policy and populate a knowledge graph that efficiently captures various inclusion and exclusion terms and rules embedded in the policy. In this paper, we describe this framework that has been built using technologies from AI, including Semantic Web, Modal/ Deontic Logic, and Natural Language Processing. We have validated our approach using industry standards proposed by the United States Federal Trade Commission (FTC) and applying it against publicly available policies of 7 cyber insurance vendors. Our system will enable cyber insurance seekers to automatically analyze various policy documents and make a well-informed decision by identifying its inclusions and exclusions.**

***Keywords*-Cyber Insurance, Ontology, Knowledge Representation, Policies.**

I. Introduction

Cybersecurity threats, like denial of service, data breaches, malware, ransomware, phishing, etc., are precipitated by hackers attacking an organization. A cyber attack on an organization can lead to data loss, reputation loss, and inhibit a business from performing its day to day activities. Often after a cyber attack, an organization has to take care of costly expenses to mitigate the negative fallout. These expenses can include costs to repair the company's reputation, cost related to technology and security upgrades, legal fees and free identity theft services to affected customers. Hence, organizations are increasingly turning to cyber insurance services to mitigate some of these costs of the potentially devastating effects after a cyber attack. Cyber insurance policies cover an organizations liabilities in case of an attack involving Personally Identifiable Information (PII) datasets, such as social security numbers, health records, credit card numbers, etc., managed by the organization.

However, in recent times, there have been situations where the insurance provided by the service provider did not include adequate coverage, or there was disagreement between the insurance service vendor and insured consumer on the coverages and exclusion rules. For instance, the breach at credit report agency, Equifax [37], led to the theft of personally sensitive information such as names, social security and credit card information, birth date and address of over 147 million consumers. Equifax reported losses in millions out of which the insurance company covered only a partial portion of the total costs. Another example is the litigation between insured consumer and insurer over a cyber incident due to NotPetya ransomware attack. The disagreement between the two parties was over the interpretation of cyber attack under the normal situation and under an act of war [38]. The insurance service company refused to pay the insurance amount arguing that the cyber incident was an act of war lead by an adversary nation and such acts of war qualified as a legitimate exclusion per their cyber insurance policy. This incident underlines the gaps and interpretation loopholes present in insurance policies currently offered by cyber insurance providers.

While businesses have migrated to the web delivery model for their services on a large scale, the cyber insurance market has not grown at the same pace. This low penetration is due to the lack of two factors: 1) Heuristically driven approach of insurance providers to competitive pricing, and 2) Competence of insurance providers to cope up with the nuances in the risks with new technology. This has resulted in a trust deficit among cyber insurance stakeholders.

One approach to address this deficit is by using semantically rich techniques to automate monitoring of cyber policies and enable smooth interaction between the cyber insurance stakeholders. We have developed a novel ontological model for the cyber insurance environment. Often heard innuendos about cyber policy confusion comes up from the complicated nature of structuring of the inclusions and exclusions in its verbatim. Romanosky et al. [8] uncovered a lack of clarity in what is covered and excluded by the policy, in the event

of a security incident. Ambiguity in designing a rule is partly responsible for lack of its comprehensibility. A comprehensive understanding of legal nuances is needed to understand the elements of a cyber insurance policy. Romanosky et al. [8] talk about how convoluted policy language leads to litigation and parties often undergo courtroom discussion to determine the validity of coverage clauses. Converting the cyber policy document from its existing textual format into a machine-processable graph database is a promising solution in such a scenario. Our vision is to build a system where, given an applicants set of requirements for cyber policy coverages, coverage limits and expected rate of coverage, the system would list matching insurance policies. A user could then choose the best policy as per their needs. Such a system would save organizations valuable resources that are currently used to manually parse through the fine prints of legal clauses in the policy. The system would help insurance seekers to identify their desired policy coverages and exclusions quickly.

In this paper, we present our novel semantically rich framework that automatically extracts essential keywords and rules from cyber insurance policy document and then populates a knowledge graph representing these extracted keywords in terms of coverages and exclusions. We have built this framework using techniques from Artifical Intelligence (AI), including Semantic Web technologies, Modal/ Deontic Logic, and Natural Language Processing (NLP). We have built our Knowledge Graph (or Ontology) using industry standards proposed by the United States Federal Trade Commission (FTC). We also present the results of the validation of this framework against publicly available policies from seven insurance providers including Hiscox [2], Chubb [3], AIG-Insurance [25], HSB [26], XL-Catlin [27], Liberty Mutual [28], and Axis Capital [29] .

The rest of the paper is organized as follows – In Section II, we discuss the related work in this area. Section III describes our framework for building and populating the knowledge graph. Section IV includes the results of our validation. We conclude in Section V.

II. Related Work

Determining the right cyber insurance policy, in line with expectations of right cost and including appropriate coverage limits, is not an easy task to achieve manually. Romanosky et al. [8], studied many popular cyber insurance policies to get an insight into how these policies are formulated, what is an insurance provider's process for premium computation, applicant security assessment, etc. They focused mainly on three areas: 1) Coverages and exclusions in the policies, 2) The security information questionnaire presented by insurance companies to the applicants; and 3) The premium computation mentioned in cyber policies. Bandopadhyay et. al. [36] also mention that proposed cyber insurance contracts tend to be overpriced because insurers are unable to anticipate customers secondary losses, resulting in poor adoption of cyber insurance by IT Managers.

A. Semantic Web

To ensure a broad understanding on the policy terms and conditions, insurers and insured organizations need to be able to exchange information, queries, and requests with some assurance that they share a common meaning. One possible approach to this issue, which we have used, is to employ Semantic Web techniques for modeling and reasoning about services related information. We have captured properties and relationships between key elements found in a cyber insurance policy into our ontological model using semantic representational languages such as Ontology Web Language (OWL) [18] and Resource Description Framework (RDF) [17]. The semantic web technologies like OWL help us in asserting various semantics from the insurance domain and represent complex hierarchies along with their domain-specific properties in a knowledge graph. On the other hand, RDF facilitates semantic interoperability and easy integration with web systems. This semantic representation allows us to perform reasoning over our knowledge graph.

Previous work has been done in populating knowledge graphs for legal text documents like service level agreements [33], web service provider privacy policy [34], cognitive assistant for legal document analytic [35], etc. All these papers have a general model where the authors develop an approach based on GATE (General Architecture for Text Engineering) [15] for automatic population of domain ontology with the information extracted from text documents. In our work, we have used a similar mechanism to process cyber insurance policy documents. Many of these systems, make use of domain-specific ontologies, which are nowadays recognized as a popular approach to represent domain knowledge. We create a cyber insurance ontology to represent various policies.

L Ma. et al. [20] explored a fusion of semantic web technologies used for data management. They argue that an ontology helped provide a structured representation of domain knowledge, as it can effectively capture various entities and their relationships. Semantic web technologies allow us to build a shared web platform for users. The author's research talks about the merits of using an ontology model for representing data and tying it up with semantic web technologies to support complex querying.

Bohme et al. [19] proposed a unifying comprehensive framework to illustrate the parameters that should be included in the model of cyber insurance. They highlight areas like, interdependent security, correlated risk, information asymmetries, cohesively and capture relationships between them. Their framework offers a unified terminology to deal with specific properties of cyber risk and helps to alleviate discovered shortcomings.

Ganino et al. [15] aimed to establish relationships between different stakeholders and cyber security components in their ontology model. Using this model, they tried to implement an understandable national cybersecurity policy framework. During the implementation, they mapped relevant aspects of the security policy to actors and functions from an ontology.

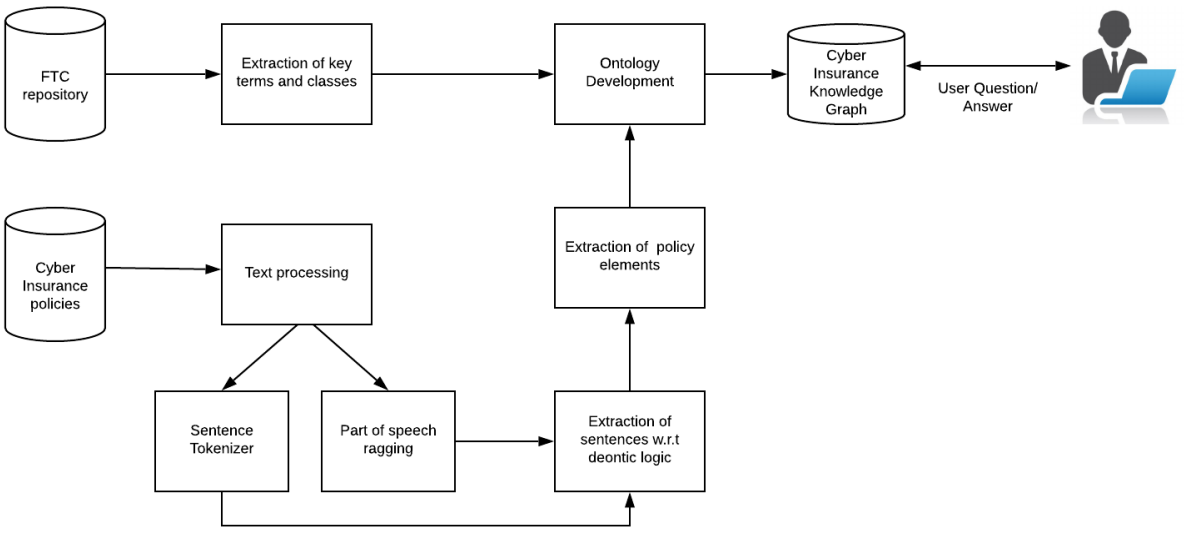

Fig. 1. System architecture of our framework.

For querying ontology, they used SPARQL query language [30]. Our web ontology for cyber insurance policy envisions a similar querying platform where we can leverage information captured in the knowledge graph and probe it using various SPARQL [30] and SWRL [31] queries.

III. Technical Approach for the Framework

In this section, we describe our approach towards creating a framework for automatic management of cyber insurance policies. Fig. 1, illustrates the overall architecture of our system. We have used Natural Language Processing (NLP) techniques and Semantic Web technologies to build this framework. As a first step, we collected cyber insurance policies of various insurance providers. We validated our knowledge graph using the work done by Romanosky et al. at the Federal Trade Commission [8]. Our system automatically extracts various coverages and exclusions from policy documents and asserts in our knowledge graph.

Our framework consists of three key parts that are described in detail in the following sections:

- **Knowledge graph for cyber insurance** We studied several publicly available insurance policies like, Hiscox [2], Chubb [3], AIG-Insurance [25], HSB [26], XL-Catlin [27], Libery Mutual [28] and Axis Capital [29]. We referred to the report generated by the Federal trade commission (FTC) document [8] to identify key classes of our cyber insurance ontology along with their relations. Section 5, describes details about the classes in ontology.
- **Automated Text Extractor for Coverages and Exclusions** This module automatically extracts various coverages and exclusions from a policy document and populates the cyber insurance knowledge graph developed in Section III-A. Our implementation consists of two components, namely, a text extraction module to extract ontology classes from cyber insurances policy documents and a core service module to populate the knowledge graph.
  We used deontic expressions, like permissions, and prohibitions, to extract policy coverages/exclusions. We expressed our deontic grammar for Permission and Prohibition as follows:
  - $<Pronoun\ |Delimiter><deontic><Noun\ phrase>$

  Examples of modal verbs used in the deontic expression for Coverages would be, “provide”, “covers”, “pay”. Modal verbs used in the deontic expression for Exclusion would be, “shall not”, “exclude”, “will not”, etc.
- **Querying & Reasoning over the knowledge graph** One of the main motive behind this work was building a cyber insurance knowledge graph that allows users to find the best possible policy that meets their needs. We host our knowledge graph on a service endpoint and users can interface with it by using semantic technologies languages like SPARQL [30].

A. Knowledge graph for cyber insurance

In this section, we describe our approach for developing and validating the cyber insurance ontology. We have developed this knowledge graph to capture the key components for insurance policies including the key terms, their definitions, the rules, and their types. Our knowledge graph, illustrated in Fig. 2, also defines the relationship between the various stakeholders in a cyber insurance environment. Having an ontological model can assist insurance providers to create a structured machine processable cyber insurance policy rules which can be automatically parsed and monitored. We designed this knowledge graph using Protege software [24] and host our knowledge graph on Apache Jena [32].

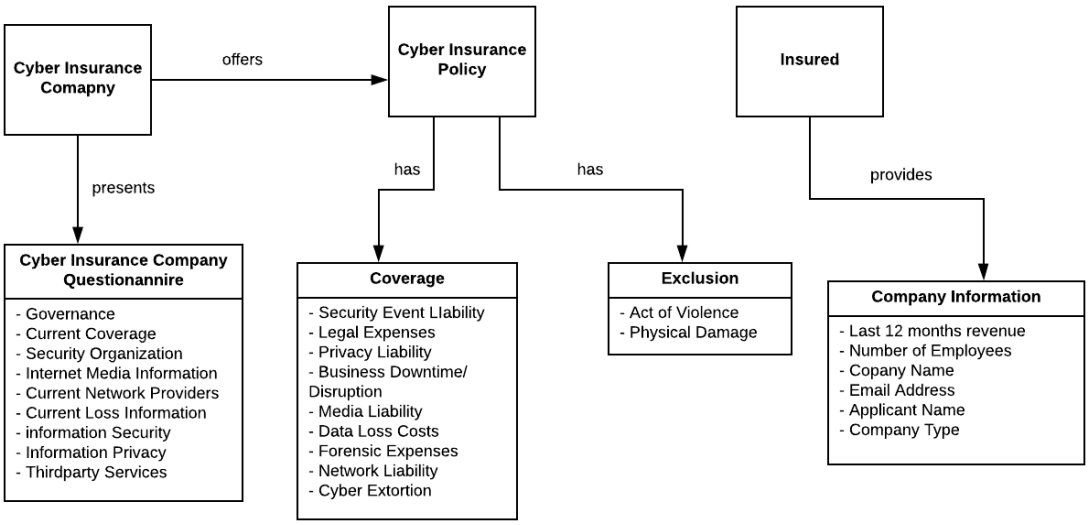

Fig. 2. Our knowledge graph represents key components in the cyber insurance policy and security questionnaire.

We studied several insurance products in details [2] [3] [25] [26] [27] [28] [29] and identified the main actors and aspects involved. Our knowledge graph captures four main classes of 'Cyber Insurance Company' representing the insurance provider, 'Insured' representing the applicant seeking cyber insurance, 'Cyber Insurance Company Questionnaire' capturing the list of questions given to the applicant to answer and 'Cyber Insurance Policy' representing the components of cyber insurance policy document. The four main classes are further detailed below:

1) **Cyber Insurance company**: This class represents a vendor offering cyber insurance policy.
2) **Cyber Insurance policy**: This class represents the actual insurance policy document. It comprises of two main sub-classes.
   - **Coverages** : These represent the elements that are covered by an insurance policy. Coverages can also be referred to as inclusions. A number of inclusions covered by an insurance policy can differ from policy to policy. The key subclasses of this class are Cyber Security Liability, Legal Expenses, Cyber Extortion, Privacy Liability, Business Downtime/Disruption, Media Liability, Network Liability, Data Loss Cost and Forensics Expenses.
   - **Exclusions** : These represent the elements that the insurance policy does not cover. For example, most of the insurance policies do not provide coverage for physical or other damages related to the act of god which are not under human control. The key sub classes of Exclusion class are 'Act of Violation' and 'Physical Damage'.

   The key subclasses of *coverages* include:
   - *Security Event Liability*: Coverage for responding to fraud or public relation expense related to the security incident.
   - *Legal Expenses*: Coverage for legal or litigation expenses in an event of security incident.
   - *Privacy Liability*: Coverage for any act of violation against personal data or sensitive Personal Identifiable Information (PII).
   - *Business Downtime*: Coverage in the event of business disruption arising due to hacking, malware or other kinds of attacks.
   - *Media Liability*: Coverage in the event of wrongful access, handling of media content or infringement of Business trade and secrets.
   - *Data Loss*: Coverage in the event of any loss to the data asset caused by DOS attack or another malicious attack on an IT infrastructure.
   - *Forensics Expenses*: Coverage for forensics expenses in the event of a security incident.
   - *Network Liability*: Coverage for any network wrongful act via hacking, the intended shutdown of the network by an operating system or unauthorized access.
3) **Cyber Insurance Questionnaire**: The insurance questionnaire includes all the details that the insurance seeker must fill out while applying for an insurance policy. The questionnaire asks questions such as details on the applicant's current security infrastructure, details about any third party services used by the applicant, security frameworks in use, etc. Once the Insurance seeker submits all the required information, it is used to formulate policy document for the company with an appropriate coverage elements and fixed exclusions at a appropriate price point. The key subclasses of the questionnaire include:

- *Third Party services*: Information on the usage of any third party services involving personally identifiable information.
- *Information Privacy*: Information concerning organization Privacy policy. It includes an assessment on the policy standards and establishes ownership of privacy policies by the organization's chief procurement officer, chief information security officer.
- *Information Security*: Information concerning security procedures established by an organization such as vulnerability scans, penetration testing, and security assessment frequency.
- *Current Coverage*: Information on current coverages sought by the organization in its existing cyber policy cover.
- *Current loss information (Loss history)*: Information on any previous security incidents and assessment of type and frequency of the attack, such as previous operating system attacks, tampering attempts etc.
- *Current Network Providers*: Information on network service providers used by an organization such as Broadband provider, Cloud services used by the organization if any, internet communication services etc.

4) **Insured / Cyber Insurance seeker**: The cyber insurance seeker is the one who seeks cyber insurance for his personal or business needs. Insurance seeker can also be referred to as an applicant or 'insured'. Insured is responsible to provide answers to the cyber insurance questionnaire presented by the insurer. This information dictates the elements of the final policy document offered by the insurer. The key subclasses of the insured class include company information about its name, address, company type, number of employees and revenue. Applicant name is the primary contact in the insured company.

Following relationship were also modeled in the ontology:

- *offers*: Cyber Insurance Company → Cyber Insurance Policy .
- *presents*: Cyber Insurance Company → Cyber Insurance Company Questionnaire.
- *has*: Cyber Insurance Policy → Coverages, Exclusions.

In our study, we observed that there is no standard format for Cyber insurance documents. There are many organizations providing cyber insurance and each insurance provider structures its policies in its own format. By capturing the key components of cyber insurance policies as a knowledge graph, we can facilitate automatic comparison of two or more policy offerings, thereby enabling consumers to make calculated choices.

B. Automated Text Extractor for coverages and exclusions

Given the complex nature of a cyber insurance policy, it is difficult to identify coverages provided and also identify exclusions applicable, in case of a security incident. The main reason for this has been the lack of clarity in the way policy language is structured. This makes it difficult for a reader to understand the policy and it also obfuscates fine legal clauses that apply to a security incident. Using ontology to represent cyber insurance policy mitigates many of these issues. An ontology provides a richer representation of various elements involved. Hence, it's necessary to develop a framework that automates the extraction of elements from a cyber insurance policy. The resulting knowledge graph enables a user to reason over it. We focus on building an automated framework for extracting coverages and exclusion from a policy document. Our extraction module uses a grammar chunking parser, which takes as input deontic grammar rules for permission (coverages) and prohibition (exclusions). Given a policy document, we use the Language Toolkit (NLTK) sentence tokenizer to get unique sentences in a policy document. We then use a word tokenizer on each of the sentences to get a list of words in the sentence. After this, the output of the tokenizer serves as the input to the Stanford part of speech (POS) tagger [12]. For example, the POS tagged output for a sentence in the policy such as We will pay Third party liability expenses will be: "('we', 'PRP'), ('will', 'MD'),('pay', 'VB')('third', 'JJ'), ('-party', 'NN'),('liability', 'NN') where PRP is a preposition, MD is a modal, VB is a verb, JJ is an adjective, and NN is a noun.

In the following sections, we describe in detail different aspects of our implementation:

Initially, for extracting coverages and exclusions from the policy document, we first tried a text based regular expression (RegEx) parser to find all the sentences in a policy that pertain to coverages and exclusions. This approach did not fare well, as regular text RegEx conforms to a narrow set of textual patterns and is less flexible. We found that using a grammar-based natural language chunking parser is better suited to solve this kind of problem. We next explored using deontic expressions on the policy documents to extract policy coverages/exclusions.

1) Defining deontic expression for extracting Coverages and Exclusions: Modal logic is a broad term used to cover various other forms of logic such as temporal logic and deontic logic [23]. Deontic logic describes statements containing permissions, and obligations, and temporal logic describes time-based requirements. Deontic logic further consists of four types of modalities:

1) Permissions / Rights: Permissions are expressions or rules that describe the rights or authorizations for an entity.
2) Obligations: Obligations expressions are the mandatory actions that an entity must perform.
3) Dispensations: Dispensations that describe optional expressions and describe non-mandatory conditions.
4) Prohibitions: Prohibitions are the expressions that specify the actions which are prohibited.

We used deontic Logic to further identify various coverages and exclusions. After tokenizing sentences in a given policy document, we further categorize them as either a coverage or

an exclusion to extract the actual coverage/exclusion classes. In order to find answers to questions like, *What all coverages this policy provides?*, we need to classify sentences into:

- *Permissions* (included coverages)
- *Prohibitions* (stated exemptions)

Whereas answering questions like, *Will my policy provide litigation cost for x?* involves extracting sentences based on:

- *Obligations* (Mandatory conditions for extending a coverage)
- *Dispensation* (Non-mandatory conditions for extending coverage)

2) Population of ontology: Once we get sentences that relate to various coverages or exclusion, the next step is to extract appropriate ontology classes.

The result of our deontic grammar parser partitions the sentence into a *subject*, *predicate*, and *object* (noun phrase chunk). The subject represents an insurance provider entity, predicate represents a deontic modal verb for coverages/exclusion, and object represents the sentence chunk representing various coverage elements. Various sentence chunks generated need to be mapped to appropriate coverage/exclusion classes defined in our ontology. We pre-process these chunk phrases and map them to one of the applicable ontology classes. For example, in a policy document a coverage clause like, "The policy covers cyber extortion damages" and a similar clause in another policy, "We will pay the costs you incur subject to cyber breach", need to be mapped to a coverage class called "Cyber Extortion." Table 3, lists major keywords generally used in various coverages. After all the coverage and exclusion sentences from the policy document are mapped to their respective ontology classes, we compile this list in the form of a payload which is sent to an ontology service. The ontology service uses an open source semantic web framework for Java (JenaApi) [32] and populates our ontology by creating class individuals. Once the ontology is fully populated, it is ready to accept a user query. For the initial phase, our ontology service stores serialized Resource Description Framework (RDF) [17] triples in-memory.

C. Querying & Reasoning over the knowledge graph

Building a web-based cyber insurance platform is one of the driving motivations for choosing an ontological way of representing cyber insurance policies. We envision that this platform would provide a way for a user to specify his set of requirements (expected coverages). Then the system would return policy matching most of the requirements. Fig. 3, illustrates a snapshot of our ontological model, depicting classes and relationships between various entities. A brief walk through of the flow of the system is as follows:

- *Insured* (represents an applicant): The user states requirement for 'Media Liability' and 'Network Liability' as required coverages for his policy.
- *Web Insurance Company* (Insurance provider): Each insurance provider *presents* their cyber policy which is automatically parsed by our system and represented by class 'Cyber Insurance Policy'. Each of the Insurance policy *fulfills* a certain set of requirements (coverages). For example in Fig. 3, Hiscox [2] policy covers Physical loss, Network liability, and Media liability, while, Chubb [3] policy just covers Media liability.
- *Relationships modeled in an ontology are*:
  - Asks for: Insured → Requirements
  - presents: Insurance provider → Cyber Insurance Policy
  - fulfills: Cyber Insurance Policy → Requirements

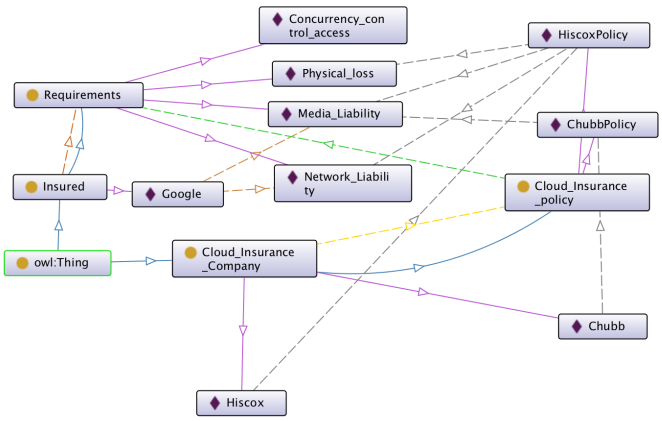

Fig. 3. Negotiation system for cyber insurance ontology.

Now if the user requires 'Media Liability' and 'Network Liability' in the insurance policy, from our system will reason over a set of rules and suggest the appropriate policy to the user. In Fig. 3, since Hiscox [2] policy fulfills both the coverages, it will be proposed as the best suitable policy for this user. In our system, we do this reasoning using SPARQL and SWRL queries.

IV. Results and Validation

A. Validating our cyber policy ontology

We used the Federal Trade Commission (FTC) document by Romanosky et al. [8] to validate our cyber policy ontology. Romanosky et al. studied over 180 cyber insurance policies. The three key components carefully examined by Romanosky et al. were the inclusions and exclusions in various cyber policies, the security questionnaire used to determine the security preparedness of an organization, and the premium computation models. We validated all our coverage and exclusion classes designed in our cyber policy ontology against the ones identified in their research. We also validated our ontology by examining seven other cyber policy documents.

B. Extracting coverage/exclusion sentences using deontic expressions

We extracted the content from cyber insurance policies that satisfied the deontic expression linguistic structure for various permissions and prohibitions. Table 1, shows key modal verbs

used in deontic grammar. Sentences with modal verbs such as 'will incur', 'will pay', 'will cover', 'be liable' should be categorized as permissions and sentences with modal verbs such as 'exclude', 'not provide', 'not incur', 'not liable' should be categorized as prohibitions. Our focus in this work is, to extract coverages from a cyber insurance policy document. Examples of sentences extracted with deontic grammar include:

- *Permission*: "We will pay cyber extortion damages and Cyber Extortion Expenses, by reason of a cyber-extortion event taking place after the Retroactive Date and prior to the end of the Policy Period" [4].
- *Prohibition*: "We shall not be liable for Damages or Expenses on account of any Claim of alleging, based upon, arising out of or attributable to any Bodily Injury or Property Damage" [4].

Modal verbs in Deontic Expression	
Permission	Prohibitions
will Incur	Exclude
will Pay	not incur
will cover	not provide
be liable	not liable

TABLE I
DEONTIC VERBS FOR COVERAGES AND EXCLUSIONS.

Extracted Policy Coverages	
Coverage category	Extracted sentence
Privacy claim coverage	We will pay Damages and Privacy Claims Expenses by reason of a Privacy Claim first made during the Policy Period.
Network Incident coverage	We will pay Damages and Network Security Claims Expenses, by reason of a Network Security Claim first made during the Policy Period.
Media expense coverage	We will pay Media Claims Expenses and Damages by reason of a Media Claim first made during the Policy Period.
Cyber Extortion coverage	We will pay Cyber Extortion Damages and Cyber Extortion Expenses, by reason of a Cyber Extortion Event taking place after the Retroactive Date and prior to the end of the Policy Period.

TABLE II
EXTRACTED COVERAGE SENTENCES AND CLASSES.

C. Extraction of coverages and exclusions for ontology development

We found that using deontic grammar worked well for extracting coverages sentences from the policy document. However, it did not extract various policy exclusions. We experimented with one of the cyber policy document, namely, Chubb policy document "Cyber Enterprise Risk Management Insurance policy" [3]. This document mentions seven high-level coverage classes , and our deontic grammar parser was able to extract coverage sentence, about all of the coverage classes from this policy document. We were able to map 6 out of 7 ontology coverage classes for these extracted sentences, giving us an accuracy of 85 percent for populating coverages. We achieved similar results while extracting details from other policies. Table 2, shows sample coverage sentences extracted from the Chubb policy document [3]. For extracting exclusion sentence, we also briefly explored using the Naive Bayes classifier. With the sample dataset from Chubbs policy document, we saw a 55 percent accuracy in classifying exclusion sentences. Exclusions in policy are often expressed a complex legal condition which, takes effect for particular set circumstances. It may not always conform to a deontic grammar rule. We add these to our extracted set. We aim to first classify sentences in the cyber policy into permissions and prohibitions. We have designed our implementation such that, given a deontically classified output, we can further use that to map it to the appropriate bucket driven by the classes defined in the ontology. To map an extracted coverage sentence into an ontology class requires some processing of the raw text. We used a bag of word approach to finding the most suitable coverage class. Table 3, shows analogous key terms representing an ontology coverage class. We were able to extract 6 out of 9 coverage classes from the cyber policy document. We validated the results of these sections as per the Federal Trade Commission (FTC) document [8].

Ontology Class Mapping	
Coverage keywords	Analogous words
Cyber Extortion	Cyber Extortion Expenses, Cyber Extortion Damages, Cyber Extortion Event
Data loss	Data Asset Loss, Data Asset Incident, Data Loss Costs
Forensic Expenses	Forensic Costs
General Liability	General Expenses
Media Liability	Media Liability, Media Claims Expenses, Media Claims Damages
Network Liability	Network Security Liability, Network Security Claims, Network Security Expenses
Privacy Liability	Privacy Liability, Privacy Claims Expenses, Privacy Damages
Security Event Liability	Security Event Liability, Security Event Claims, Security Event Expenses
Business Downtime	Business Downtime, Business Disruption, Business Interruption Loss, Business interruption Incident

TABLE III
KEY CLASSES AND ANALOGOUS MAPPING KEY TERMS.

V. CONCLUSION

Cyber insurance policy documents are maintained currently as textual documents with legalese that is hard to comprehend. These generally have no defined formal structure and so

are not machine processable which makes them difficult to automate. We have developed a semantically rich framework to automate monitoring of cyber insurance policy documents. The system automatically extracts relevant deontic expressions, policy expressions, and other legal sentences. We have also developed a semantically rich knowledge graph to capture information about cyber insurance providers, insurance seekers, and policy elements. In this paper, we also illustrate how a user can query the knowledge graph and use it to answer various queries related to policy coverages/exclusions. It also allows the user to select the policy, that best suits her needs, among various possible cyber policies offered by different vendors.

Automatically extracting and storing essential elements of cyber insurance policy in a knowledge graph, is the first step towards building a policy negotiating system, which can negotiate with policy vendors on behalf of the user. In the future, given a set of user requirements like expected inclusions, inclusion limits, and expected rate, the system would be able to negotiate the best cyber insurance policy for a user utilizing our knowledge graph.

ACKNOWLEDGMENTS

This research was partially supported by a DoD supplement to the NSF award 1439663: NSF I/UCRC Center for Hybrid Multicore Productivity Research (CHMPR). We thank Amey Sane for providing valuable input.

REFERENCES

[1] Karttunen, L., Chanod, J.-P., Grefenstette, G., and Schiller, A. (1996). Regular expressions for language engineering. Journal of Natural Language Engineering, 2(4):305328.

[2] Hiscox PLC. (2015). Cyber and data Policy wording, WD-PIP-UK-CD(2) 13388 05/15.

[3] Chubb Insurance policy. https://www.chubb.com/cz-cz/assets/documents/chubbpp-cyber-enterprise-risk-management-en.pdf.

[4] Marthie Grobler, J.C. Jansen van Vuuren, Louise Leenen ImplementationofaCyberSecuritypolicyinSouthAfricaImplementationofaCyberSecuritypolicyinSouthAfrica.pdf

[5] James Geller, Soon Ae Chun, Arwa Wali, "A Hybrid Approach to Developing a Cyber Security Ontology", Proceedings of 3rd International Conference on Data Manage-ment Technologies and Applications, pp. 29-31, August (DATA 2014).

[6] N. Gcaza, R. V. Solms, and J. V. Vuuren. An Ontology for a National CyberSecurity Culture Environment. Proceedings of the Ninth International Symposium on Human Aspects of Information Security & Assurance (HAISA 2015), (Haisa):1 10, 2015.

[7] Amina Souag, Camille Salinesi and Isabelle Comyn-Wattiau2 Ontologies for Security Requirements: A Literature Survey and Classification

[8] S. Romanosky, L. Ablon, A. Kuehn, T. Jones, Content analysis of cyber insurance policies: How do carriers write policies and price cyber risk?, 2017.

[9] Insurance Information Institute. (n.d.-a). Insurance Industry at a Glance. III. Retrieved from http://www.iii.org/fact-statistic/industry-overview

[10] Meenachi, N.: Web ontology language editors for semantic web- a survey. International Journal of Computer Applications 53(12), 1216 (2012)

[11] Delia Rusu, Lorand Dali, Blaz Fortuna, Marko Grobelnik, and Dunja Mladenic. 2007. Triplet extraction from sentences. Proceedings of the 10th International Multiconference "Information Society – IS 2007", A:218–222, October.

[12] Stanford POS Tagger. http://nlp.stanford.edu/lexparser.shtml.

[13] Regular Expression Parser. https://dzone.com/articles/a-guide-to-parsing-algorithms-and-technology-part.

[14] Word and Sentence Tokenizer. https://training-course-material.com/training/Natural Language Processing in Python 5.2.Chinking.C2.B6.

[15] Ganino G, Lembo D, Mecella M, Scafoglieri F. Ontology population for open-source intelligence: A GATE-based solution. Softw Pract Exper. 2018;129. https://doi.org/10.1002/spe.2640

[16] Jansen van Vuuren J, Leenen L, Zaaiman J. Using an ontology as a model for the implementation of the national cybersecurity policy framework for South Africa

[17] O. Lassila, R. Swick, Resource Description Framework (RDF) Model and Syntax Specification, Feb. 1999.

[18] D.L. McGuinness, F. van Harmelen, "OWL Web Ontology Language Overview", Feb. 2004. Consortium, 2004.

[19] Bohme, R. and Schwartz, G. 2010. Modeling cyber-insurance: Towards a unifying framework. In Proceedings of the Workshop on the Economics of Information Security (WEIS).

[20] L. Ma, J. Mei, Y. Pan, K. Kulkarni, A. Fokoue, and A. Ranganathan, "Semantic web technologies and data management," in Proc. of W3C Workshop on RDF Access to Relational Databases, 2007.

[21] Oltramari, A., Cranor, L.F, Walls, R., McDaniel, P., "Building an Ontology of Cyber Security", in STIDS 2014 (9th International Conference on Semantic Technology for Intelligence, Defense, and Security, 2014

[22] Oltramari, Alessandro, Diane S. Henshel, Mariana Cains and Blaine Hoffman. Towards a Human Factors Ontology for Cyber Security. STIDS (2015).

[23] Modal logic. Modal Logic: http://plato.stanford.edu/entries/logic-modal/

[24] Musen, M.A. The Protege project: A look back and a look forward. AI Matters. Association of Computing Machinery Specific Interest Group in Artificial Intelligence, 1(4), June 2015. DOI: 10.1145/2557001.25757003.

[25] AIG Insurance policy. http://www.aig.com/content/dam/aig/america-canada/us/documents/business/cyber/ cyberedge-cyber-liability-insurance-brochure.pdf

[26] HSB Insurance policy. https://www.constructaquote.com/media/1519/hsbeil_cyber_policy_wording.pdf

[27] XL Catlin Insurance policy. https://xlcatlin.com/-/media/xlinsurance/pdfs/professional/cyber-liability/xl-catlin_ifl_apac_cyber-product-sheet.pdf

[28] Liberty Mutual Insurance policy. https://www.libertyspecialtymarkets.com/wp-content/uploads/2015/01/LSM074_Product_overview-Cyber_risksSCREEN.pdf

[29] Axis capital Insurance policy. https://www.axiscapital.com/docs/default-source/docs/insurance/us/professional-lines/axis-pro-mpl-solutions-insurance-policy-mpl-1100-(11-09)-specimen.pdf?sfvrsn=5b2c0b2f_0

[30] Prud'hommeaux, E. and Seaborne, A. 2008. SPARQL query language for RDF. W3C recommendation. http://www.w3.org/TR/rdf-sparql-query/.

[31] OConnor, M., & Das, A. (2009). SQWRL: A query language for OWL. Proceedings of OWL: Experiences and Directions (OWLED), the fifth International Workshop.

[32] Apache Jena API. https://jena.apache.org/documentation/inference/

[33] Aditi Gupta, Sudip Mittal, Karuna P Joshi, Claudia Pearce and Anupam Joshi. Streamlining management of multiple cloud services. 2016 IEEE 9th International Conference on Cloud Computing (CLOUD).

[34] Karuna P Joshi, Aditi Gupta, Sudip Mittal, Claudia Pearce, Anupam Joshi and Tim Finin. Semantic approach to automating management of big data privacy policies. 2016 IEEE International Conference on Big Data (Big Data).

[35] Karuna P Joshi, Aditi Gupta, Sudip Mittap, Claudia Pearce, Tim Finin and others. Cognitive assistant for legal document analytics. 2016 AAAI Fall Symposium Series.

[36] Bandyopadhyay, Tridib, Vijay S. Mookerjee, and Ram C. Bao. Why IT managers don't go for cyber-insurance products, Communications of the ACM - Scratch Programming for All 52, no. 11 (2009): 68-73.

[37] Equifax Breach. https://www.insurancejournal.com/news/national/2018/03/04/482301.htm

[38] NotPetya Effect. https://www.schneier.com/blog/archives/2019/02/cyberinsurance_.html