Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 1990–1999
Marseille, 11–16 May 2020

Semantic Annotation for Improved Safety in Construction Work

Paul Thompson[1], Tim Yates[2], Emrah Inan[1] and Sophia Ananiadou[1]
[1]National Centre for Text Mining, Department of Computer Science, University of Manchester, UK,
[2]Health & Safety Executive, HSE Science & Research Centre, Buxton, UK
{paul.thompson,emrah.inan,sophia.ananiadou}@manchester.ac.uk, tim.yates@hse.gov.uk

Abstract

Risk management is a vital activity to ensure employee safety in construction projects. Various documents provide important supporting evidence, including details of previous incidents, consequences and mitigation strategies. Potential hazards may depend on a complex set of project-specific attributes, including activities undertaken, location, equipment used, etc. However, finding evidence about previous projects with similar attributes can be problematic, since information about risks and mitigations is usually hidden within and may be dispersed across a range of different free text documents. Automatic named entity recognition (NER), which identifies mentions of concepts in free text documents, is the first stage in structuring knowledge contained within them. While developing NER methods generally relies on annotated corpora, we are not aware of any such corpus targeted at concepts relevant to construction safety. In response, we have designed a novel named entity annotation scheme and associated guidelines for this domain, which covers hazards, consequences, mitigation strategies and project attributes. Four health and safety experts used the guidelines to annotate a total of 600 sentences from accident reports; an average inter-annotator agreement rate of 0.79 F-Score shows that our work constitutes an important first step towards developing tools for detailed semantic analysis of construction safety documents.

Keywords: Semantic Annotation, Named Entities, Construction Safety

1. Introduction

The health and safety of workers is of paramount importance in any construction project. As such, risk management is a vital activity, whose aims are to identify the potential risks to workers at each stage of the project and to determine appropriate safety measures that can mitigate these risks as effectively as possible.

An important source of evidence for risk management comes from existing document repositories. Reports on construction-related injuries and illnesses, site inspections and prosecutions, along with health and safety guidelines, can collectively provide details about which accidents have occurred during previous projects, together with possible reasons (e.g., failure to put adequate protection measures in place) and recommendations of how incidents can be minimised or avoided in future.

Such document repositories usually contain a large number of (possibly lengthy) documents. However, information about previous projects is only likely to be relevant to risk management for new projects if there is a match between at least some of their attributes. These attributes may include location type (e.g., the terrain or type of construction), construction activities undertaken (e.g., demolition, digging), equipment used (e.g., ladders, drills), etc. Furthermore, information about incidents, potential reasons and possible mitigations may be fragmented both within and across a range of different documents. As such, the processes of finding relevant supporting documents, locating important information that may be deeply buried within them, and identifying links between information from different documents can be extremely laborious and time-consuming. There is thus a high chance of missing potentially important information that could make the difference between the life or death of a worker.

In a range of domains, text mining (TM) methods have been applied to automatically analyse the content of free text documents, in order to identify various aspects of their structure and meaning. Named Entity Recognition (NER) is a fundamental step in TM, whose aim is to detect and semantically classify mentions of important domain-specific concepts. In our case, these would include types of accidents, features of building sites, equipment or tools and safety measures, etc.

Recognised named entities (NEs) form the basis for the application of more sophisticated relation detection methods, which extract precise structured representations of the knowledge hidden in documents, by discovering which entities are related to each other, and how. In construction safety, examples may include linking accidents to relevant contextual factors (e.g., equipment being used when they occurred) or linking specific safety measures to the hazards that they aim to alleviate.

Within the construction safety domain, NER is a largely unexplored problem, mainly due to the lack of suitably annotated corpora, which are required for the development and evaluation of supervised machine learning NER methods. In response, we have defined a novel NE annotation scheme for application to construction safety documents, accompanied by detailed guidelines to ensure consistent labelling by annotators. The design of the scheme (a joint effort between experts in health and safety and corpus development) was guided by reference both to existing models of construction safety knowledge and to a range of relevant document types. An iterative cycle of experimental NE labelling and discussion between the experts allowed us to converge towards a scheme that encodes a range of important concepts, including hazards, consequences, mitigation strategies and project attributes, while also being feasible to apply by annotators in a reliable and consistent manner.

We subsequently trained four health and safety experts in the application of the scheme, and asked them to label a total of 600 sentences from accident reports. An overall inter-annotator agreement rate of 0.79 F-score, together with the finding that most categories in our scheme were frequently annotated in text, help to verify the validity and suitability of our scheme. As such, the scheme and corpus will act as an important stimulus for research that moves beyond document-level classification methods (which have formed the bulk of previous research in this area), to more sophisticated methods that can perform more detailed semantic analyses of the content of these documents. The corpus may be downloaded from `http://www.nactem.ac.uk/discovering_safety/`

2. Related Work

Most text mining studies on documents relating to the construction industry and/or accidents have focussed on document-level methods. For example, semantic similarity techniques have been applied to organise construction project documents into semantic clusters (Al Qady and Kandil, 2014). Similar methods have been used to aid with case-based reasoning, in which the retrieval of documents describing cases that are similar to a situation of interest, such as dispute resolution in construction accidents (Fan and Li, 2013) or mitigating risks in construction projects (Zou et al., 2017), can help to provide evidence about how best to deal with the situation. Approaches to the automatic classification of documents have used pre-defined topics from a construction information classification system (Caldas and Soibelman, 2003) or different categorisations of injuries, incidents or hazards (Bertke et al., 2012; Taylor et al., 2014; Goh and Ubeynarayana, 2017) as their classification basis. Chi et al. (2014) classify documents from multiple databases according to both constructions activities and hazards, with the aim of linking documents that mention unsafe scenarios with those mentioning safe approaches to the same construction activities.

More detailed analyses of the semantic content of construction documents have mainly been based on rules, syntactic structure and/or the use of ontologies. Al Qady and Kandil (2009) apply shallow parsing to construction contract documents to identify active and passive concepts, together with relations between them, while Zhang et al. (2019) enrich document-level accident classification with automatic extraction of phrases denoting the causes of the accidents from document titles, using syntactic patterns.

Ontologies, which provide inventories of the concepts used within a domain, can drive methods that carry out more detailed concept-level annotation in text, by automatically marking-up phrases that correspond to mentions of different ontology classes. The rule-based approach described in Zhang et al. (2013) combines syntactic information with semantic information from dictionaries and an ontology to extract concepts and relations from construction regulatory documents. BIMTag (Gao et al., 2017) is a semantically enhanced search system that makes use of the domain-specific Industry Foundation Classes (IFC) ontology (ISO, 2013) to semantically annotate BIM product documents with phrases corresponding to physical objects used within construction, e.g., walls, beams and doors.

However, an issue with many ontologies is that they are not well–suited to supporting accurate and comprehensive NER. A given concept can often be described in text using a range of different synonyms or related terms, but it is unlikely that these will be listed exhaustively in the ontology. For example, while the IFC lists natural language names for concepts in a number of different languages, only a single name per language is provided. Both Gao et al. (2017) and Zou et al. (2017) try to overcome such issues by using the comprehensive lexical database of general language, WordNet (Miller, 1995), to obtain additional synonyms. However, acknowledging that general language resources may not sufficiently account for domain-specific language usage, Zou et al. (2017) expand an initial list of risk management terms by applying the Word2Vec algorithm (Mikolov et al., 2013) to a collection of risk cases and to Wikipedia, and collect related terms from the obtained word embeddings. Word2Vec has also been applied to an 11 million word corpus of construction-related text (Tixier et al., 2016b); the resulting vectors provide further scope to discover semantically related terms within the domain. Chen and Luo (2019) describe a further domain-specific application of word embeddings, i.e., to create knowledge graphs that encode the semantic relatedness between terms used in different types of sentences in literature abstracts that relate to construction management.

A further body of work (Desvignes, 2014; Esmaeili and Hallowell, 2012; Prades, 2014) considers that construction situations can be comprehensively characterised by a finite number of observable fundamental construction site attributes. Eighty distinct concepts corresponding to injury precursors are identified, including equipment (*hammer, forklift*), materials (*concrete, cable*), working environment (*uneven working surface, adverse low temperatures*) and actions/activities (*welding, stripping*). Tixier et al. (2016a) used this as the basis to develop a tool for automated content analysis of construction injury reports. Recognition of the mentions of the precursors is augmented by detection of seven injury types, five categories of injured body parts and nine energy sources. A complex system of manually constructed dictionaries and rules is used to detect various ways in which the attributes could be described in text, with an overall accuracy of 95%. The system was subsequently applied to 5298 accident reports (Tixier et al., 2017), followed by unsupervised data mining techniques, to discover attribute combinations that most frequently contribute to injuries.

3. Our Approach

Compared to previous work, our annotation scheme targets the development of supervised learning methods that can recognise mentions of a wider range of concepts relevant to construction safety, and can categorise them in a finer-grained manner. Thus, in contrast to Gao et al. (2017), NER will not be reliant on an ontology that lacks sufficient coverage of domain-specific synonyms, nor will it be restricted to recognising a fixed set of concepts, as was the case in Tixier et al. (2016a). Although the latter framework could theoretically be extended, they acknowledge that this would require involvement of trained content analysts and re-calibration of the tool.

Creating annotated corpora does, nevertheless, bring its own challenges. Manual annotation is laborious and time-consuming, and achieving consistency amongst annotations

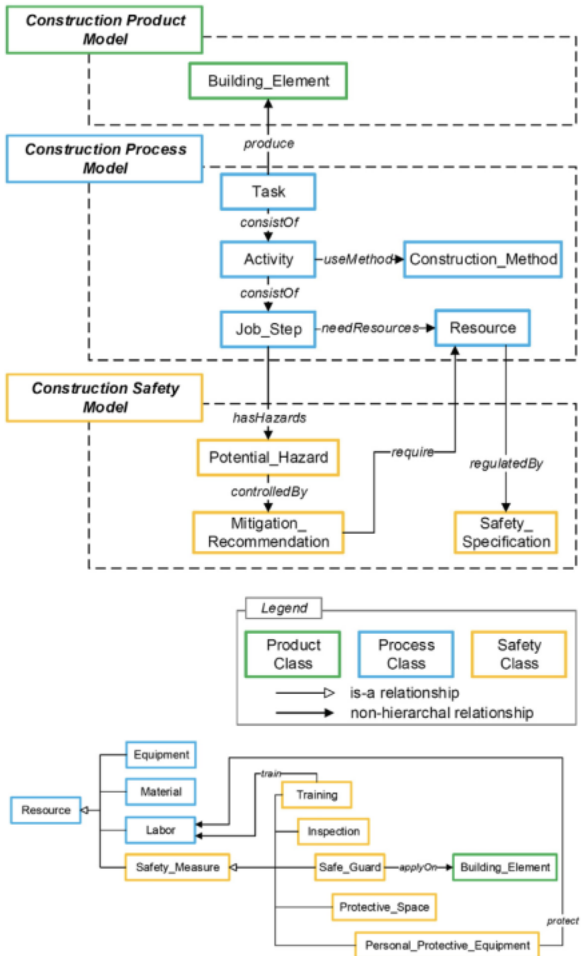
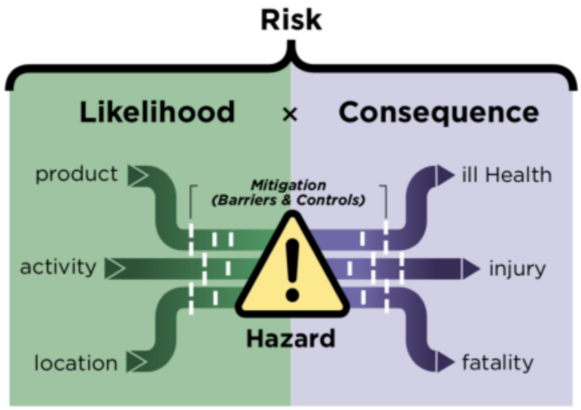


Figure 1: Part of the Construction Safety Ontology (Zhang et al., 2015)



©image courtesy of Arup and AEC3

Figure 2: Modelling of hazards, sources and consequences in PAS 1192-6 (BSI, 2018)

is vital if the corpus is to be useful for machine learning purposes. Such consistency can usually only be attained by creating detailed annotation guidelines and training annotators. However, once constructed, annotated corpora are a valuable resource that can be reused in training and evaluating different tools and algorithms.

We address some of the problems of manual corpus construction by using APLenty, an easy-to-use, web-based annotation system (Nghiem and Ananiadou, 2018), which aims to maximise both the efficiency and effectiveness of annotation. Firstly, incorporation of active learning (Settles, 2009) means that annotators are only asked to label the most informative and representative examples, rather than the entire corpus, thus reducing the time required. Secondly, the use of proactive learning algorithms (Li et al., 2017) helps to maximise annotation quality by automatically determining which annotator is likely to label an instance most accurately, based on their previous performance.

4. Annotation Scheme Design

4.1. Structured Knowledge Representations

Ontologies and classification systems constitute models of domain-specific knowledge. They define fundamental concept types, and specify how these concepts are linked via various types of relations to encode different types of knowledge. Automatically extracting and structuring knowledge from text according to such models can enhance effective sharing and reuse of information hidden in text. Potential uses of this structured information in the construction domain include integrating safety information within Building Information Modelling (BIM) software (a widely used approach to design and construction, using a digital representation of the building process to facilitate the exchange and interoperability of information), coding of occupational injury and illness incidents (U.S. Department of Labor Bureau of Labor Statistics, 1997), or the (semi) automatic creation of risk registers (Figure 3 shows a risk register format recommended by the British Standards Institution (BSI, 2018)). We aim to develop a scheme for NE annotation that is compatible with concepts in existing domain specific ontologies, e.g. (Guo and Goh, 2017; Zhang et al., 2015), while at the same time being able to support a range of practical tasks, such as those outlined above.

Figure 1 provides an overview of the Construction Safety Ontology proposed by Zhang et al. (2015), which is compatible with BIM. A construction *task* or activity produces a *building element*, e.g., a roof or floor. Specific *job steps* that may lead to hazards are further broken down according to the *equipment* and *materials* required. Each activity may be associated with *hazards*, which can be controlled through *mitigation recommendations*. These require specific *safety measures*, which are further split into *Training, Inspection, Safe Guards* and *Personal Protective Equipment*.

In the BSI's recent PAS 1192-6 recommendations (BSI, 2018) about structuring health and safety knowledge in BIM (see Figure 2), *activities* are also seen as sources of hazards, as are specific *products* (parts of structures, prefabrications, materials and substances) and *locations* of work. This model also incorporates the specific *consequences* of a hazard (injuries, ill health and fatalities); the identification of such consequences can be important in compiling structured risk registers of the type recommended in PAS 1192-6 (Figure 3), which assess the level of severity of risk consequences. A more detailed analysis of such consequences is suggested in the Occupational Injury and Illness Classification Manual (U.S. Department of Labor Bureau of Labor Statistics, 1997), where consequences are broken down according to the *Nature of Injury or Illness* (e.g., *fractures*)

Risk Name	Hazard Category	Risk Description	Associated Product	Associated Activity	Associated Location	Agreed Mitigation	Risk Likelihood	Risk Consequence	Level Of Risk	Risk Documentation	Date Raised
AAM05	Falls	Falling from height / Damage to building	roof light	cleaning and maintenance	atrium	A detail cleaning and maintenance strategy for the building is being developed with specialist consultant. The atrium roof will be designed to be walked on and appropriate access, drainage and slip resistance will be considered.	High	Very High	Very High		2013/05/14
AAM03	Falls	Falling from height; Items falling from height	internal and external façade glazing	replacement	cut back areas of the building	A detail cleaning and maintenance strategy for the building is being developed with specialist consultant.	Low	High	Moderate		2013/05/14
AAM06	Falls	Falling from height / Damage to building	glazing, feature lighting	cleaning and maintenance	roof, bridges,	A detail cleaning and maintenance strategy for the building is being developed with specialist consultant. A travelling beam and demountable cleaning cradle are being considered to allow safe access to all areas.	High	Moderate	Moderate		2013/05/14

Figure 3: Example of a structured risk register, taken from PAS 1192-6 (BSI, 2018)

and the *Part of Body Affected*.

Based on the analysis above, we can conclude that phrases representing the following types of information are important in modelling construction safety knowledge:

- Parts of structures
- Construction work activities
- Building materials
- Equipment used in construction
- Locations, spaces, or features of a building site that characterise the work environment
- Risks, hazards and incidents
- Incident consequences, e.g., specific injuries or illnesses
- Safety measures that can mitigate risks

4.2. Experimental Annotation Phase

With the above in mind, we gathered a small corpus of document extracts from the construction sector repository of the Health and Safety Executive (HSE), which includes incident reports, enforcement activity documents and safety guidance. An iterative process of experimental labelling, discussion and refinement between two of the authors (one with health and safety experience and the other with annotated corpus development experience) aimed to explore whether and how the types of concepts outlined above manifest themselves in a range of relevant document types. The purpose was to define a scheme that can be used to markup potentially complementary types of information in documents from range of different textual sources. We tried to achieve an appropriate trade-off between domain-specific knowledge that is desirable to encode, and what can feasibly be labelled by annotators in a reliable and consistent manner, without having to make too many difficult choices. Important decisions made during this process included the following:

1. A general *Protection_Measure* category was selected to label all safety and protection measures. We found that the finer-grained categories of Zhang et al. (2015) were not exhaustive of all types of measures found in text; trying to enumerate additional sub-types (and expecting annotators to differentiate between them) is considered too complex.
2. Following the guidelines provided by the U.S. Department of Labor Bureau of Labor Statistics (1997), we separate hazard consequences into *Body_Part_Injured* and *Harmful_Consequence* (*break, sprain*, etc.). This type of classification has facilitated analysis of occupational accidents in various work sectors, including construction (Glazner et al., 2005), and could also be helpful for tasks such as assessing consequence severity in a risk register.
3. Based on the knowledge representations outlined in Section 4.1, we originally tried to separate the concept types of *structure, materials* and *location*. However, we found this distinction to be difficult in practice, since the choice may depend on subtle contextual cues and/or on how the situation is viewed. For example, a roof may be viewed as the *location* of a hazard or part of *structure*, while a flooring slab could be considered part of a *structure* if it is already in place, but as a *material* if is being laid. We thus decided to use a single, overarching category, *Physical_Environment*, to cover the three concept types. The exact contribution of such entities to the description of events (e.g., whether they constitute the location of a hazard or a material being used for construction) may be more precisely determined by relation extraction methods.
4. Descriptions of hazardous events may be complex, often comprising a number of unexpected events that lead to the occurrence of an incident. This is exemplified in the following sentence, where the jamming of the section of steel and its subsequent springing eventually cause the operator to be struck: *A section of steel* ***jammed*** *on the input trestles and when trying to release it the section* ***sprung*** *towards the operator* ***striking*** *his right arm.* Accordingly, words/phrases denoting both the eventual incident and its precursor events should be annotated, to allow the identification the types of events that can lead to a "loss of control".

4.3. Final Scheme and Guidelines

In Table 1, we summarise our finalised entity annotation scheme, which includes seven distinct concept types. The table provides brief descriptions of each entity type, along with a number of illustrative examples. A detailed set of guidelines was also created to assist annotators in the accurate application of the scheme. For each category, the guidelines are organised under two main headings, which aim to maximise annotation consistency:

- **Scope**: the semantic scope of the words/phrases that should be considered for annotation
- **Span**: the exact extent of text that should be marked up when creating an annotation

The guidelines use a range of examples to characterise more precisely what should be *included* and/or *excluded* from the annotation scopes and spans. As an example, the **scope** of the *Construction_Activity* category excludes verbs such as *using*, which do not provide meaningful information about the nature of the activity being carried out. Meanwhile, adverbs are to be included in the annotation **span** of *Construction_Activity* annotations in cases where they indicate the direction of movement, e.g. *climb up*, since directional information could be important in determining the precise nature of activities that can lead to the occurrence of incidents/accidents/hazardous events.

5. Annotation Using the Scheme

Four annotators with a health and safety background were trained in the application of the scheme using the guidelines. They then used the APLenty system (Nghiem and Ananiadou, 2018) to label a total of 600 sentences from HSE's repository of desensitised Reporting of Injuries, Diseases and Dangerous Occurrences Regulations (RIDDOR) reports, which UK law requires employers to complete when certain types of workplace accidents occur. The open source status of these reports allows us to make the annotated corpus freely available for research. Two-thirds of the sentences were labelled by at least two annotators, allowing calculation of inter-annotator agreement (IAA).

Since the primary purpose of RIDDORs reports is to report injuries, mentions of protection measures are rare. Accordingly, we decided not to annotate *Protection_Measure* in this initial corpus. However, our preliminary analysis of other types of documents, i.e., enforcement activity documents and safety guidance, strongly suggests that this category will be useful when the scheme more widely applied.

5.1. Inter-Annotator Agreement

We calculated IAA in a pairwise fashion, and show the average agreement rates over the five pairs of annotators with overlapping annotations in Table 2. The widely used Cohen's kappa is not suitable for calculating agreement here, because it requires the total number of annotated items to be known in advance. Hence, we followed a number of other related efforts (Hripcsak and Rothschild, 2005; Ju et al., 2019; Thompson et al., 2018) by calculating IAA in terms of F-score. We show figures for both *exact span matching* (cases where both annotators selected exactly the same NE category and the same text span) and *relaxed span matching* (cases where annotators selected the same NE category and overlapping text spans). Although consistent span is preferable for training machine learning algorithms, the potential variability in textual expression of certain NEs can make many span selection choices difficult; relaxed matching statistics provide evidence about the extent to which annotators agree on which NEs to annotate, even when annotated spans do not match exactly.

Five out of the six categories annotated achieve relaxed match F-scores of 0.75 and above, showing that, in most cases, annotators have a good common understanding about which phrases constitute NEs and how they should be categorised. An exception to this is *Construction_Activity*, which appears to be slightly more problematic to identify, as do the precise span boundaries of certain entities.

5.2. Discrepancy Analysis

We outline some recurring types of annotator discrepancies in Table 3. Several types of span disagreements concern situations that are already addressed in the guidelines, such as the inclusion of structural materials in *Physical_Environment* spans (e.g., ***oak*** *staircase*) or that annotations of certain categories may span over noun phrases and their following prepositional phrases, when this is necessary for complete characterisation of the concept (e.g., ***foot of the*** *stairs*). In some cases, these specific guidelines seem to be disregarded by annotators in favour of the more general rule that spans should be kept as short as possible. We plan to experiment with enhancements to the guidelines (e.g., a brief summary of the most important aspects and/or a flow chart) that would make it easier for annotators to make more consistent decisions.

Other span discrepancies may result from decisions that can be difficult to make without a linguistic background; enrichment of the guidelines with further illustrative examples may help. For example, the guidelines state that adverbs that follow verbs should be annotated when an action/activity can only be correctly understood in conjunction with the adverb. However, determining this could depend on potentially subtle semantic issues, e.g., since the action of "raising" is intrinsic to the meaning of *hoist*, the inclusion of the adverb *up* may be considered redundant. In contrast, *slip off* denotes a more specific event than the verb *slip* alone. A further guideline states that annotations corresponding to actions should only span a verb *and* its object if they collectively constitute a distinct unit of meaning, e.g., *lost his footing*, which is very different to the meaning of *lost* alone. The verb *catch* also has a number of possible senses, which can depend on the nature of the object, e.g., a body part object (*caught his foot*) "selects" the meaning of "becoming entangled". However, there is still a difference between this and the more idiomatic *lost his footing*; *caught* would have the same interpretation whichever body part is chosen, and this interpretation is probably the default one in a construction context.

Certain annotation scope definitions may also need more precise definitions. This is especially the case for *Construction_Activity*, which exhibits the lowest relaxed matching rate. We found that "core" construction activities, like *drilling, dismantling, lifting* or *offloading* generally achieve high levels of agreement. However, other "supporting" ac-

Category	Description	Examples
Protection_measure	Measure or action to increase safety at work, prevent hazards and/or reduce the risk of accidents. Includes: • General activities to increase safety at work • Protective clothing • Equipment, mechanisms and devices aimed at increasing safety	• *We have reviewed the* ***[risk assessment]*** *and* ***[safe system of work]*** *for this task.* • *Once on the verge one of the operatives came across and undid his left* ***[lace up boot]*** • *We are also looking at retro fitting an* ***[auto hose feed device]***
Body_part_injured	Part of the body injured during a construction-related incident while carrying out construction-related activity	• *He landed heavily on his* ***[elbow]*** • *The large roll of protector fell from the worktop onto the IP 's* ***[right wrist]****.*
Harmful_consequence	A change in a person's physical state or wellbeing that is a consequence of a harmful incident or a health and safety issue	• *When standing up he felt an* ***[ache]*** *in his lower back and could not continue .* • *Kneeling on concrete while using float to spread cement caused* ***[alkali burns]*** *through PPE to knees and shins.*
Construction_activity	Physical activity or action carried out as part of everyday tasks relating to construction work; excludes unintended actions/accidents	• *They had* ***[loaded]*** *the goods hoist board by board from a stack at ground level.* • *Whilst* ***[exiting]*** *the loft, using a TItan class 1 ladder, the injured person lost his footing whilst* ***[descending]*** *the ladder*
Equipment	Equipment used during the course of carrying out construction activities. Includes: • Mechanical equipment and vehicles • Equipment used for access to construction sites • Construction tools (hand or mechanical)	• *The Injured person (IP) was working on the* ***[crash cushion vehicle]****.* • *IP fell while exiting a plot (18-19) onto a* ***[scaffolding platform]****.* • *He was drilling with a* ***[hand held battery operated drill]***
Physical_environment	Phrase referring to the physical environment of the construction work. Includes: • Permanent or semi-permanent structures (including parts of such structures and materials used to build them) that will stay in place once a construction project has finished • Construction sites • Natural environments/locations where construction is taking place	• *IP was putting protection film onto* ***[work surfaces]*** *in the tenant 's* ***[kitchen]****.* • *They then pushed the loaded trolley across the* ***[floor slab]*** *and were in the process of pushing the trolley up a small* ***[ramp]*** • *The injury took place on a* ***[construction site]****.* • *The ladder was positioned on the* ***[grass lawn]****.* • *The ramp was made from a 2440 mm x 1220mm 18mm thick* ***[plywood sheet]*** *and screwed to the* ***[timber supports]*** *underneath.*
Hazard	Unexpected/unintentional and (potentially) harmful accident or incident occurring during construction activities. Includes both: • Words/phrases denoting events which cause (or may cause) harm • Words/phrases denoting unintentional events that are precursors to the (potentially) harmful event.	• *A section of steel* ***[jammed]*** *on the input trestles and when trying to release it the section* ***[sprung]*** *towards the operator* ***[striking]*** *his right arm* • *The operative* ***[touched]*** *his head on a fire alarm and received an* ***[electric shock]****.* • *During modification of scaffold, the scaffolder has over stretched and has* ***[lost footing]*** *on wet boards causing him to* ***[fall backwards]****.*

Table 1: Summary of named entity annotation scheme for construction safety

Category	Exact Span Match (F-Score)	Relaxed Span Match (F-Score)
Body_part_injured	0.81	0.92
Harmful_consequence	0.72	0.78
Construction_activity	0.56	0.68
Equipment	0.72	0.84
Physical_environment	0.61	0.75
Hazard	0.57	0.79
Total	**0.66**	**0.79**

Table 2: Inter-annotator agreement

tions/activities that form part of these main tasks, or which are undertaken to prepare for them, are annotated less consistently. Examples include activities like *passing* something from one person to another or *twisting* to fit into a space. We consider that it is correct to annotate such actions, given that hazards may equally be present when they are being undertaken.

While high agreement for *Equipment* is generally achieved, we noticed that some disagreements concerned equipment that is not directly used to carry out construction work, e.g., *lights*. Nevertheless, such entities are still considered relevant, given their potential to contribute towards an incident, e.g., *He caught his foot on the lead running to a* ***plasterer's light***. Similarly, component parts of equipment may also be relevant, even though we did not mention them explicitly in the initial guidelines, e.g. a loose *rung* of a ladder could cause a fall. Several scope issues with *Haz-*

Category	Span disagreements	Scope disagreements
Body_part_injured	• Specificity: ***[tendon of]*** *thumb of left hand*	• Lesser-known body parts: *ulna*
Harmful_consequence	• Preposition: *cut* ***[through]*** • Severity marker: ***[extremely]*** *painful*	• Non-specific locations: *internal injuries* • Minor consequences: *brushing*
Construction_activity	• Preposition: *hoist* ***[up]*** • Verb object: *drill* ***[a hole]***	• Non-core construction activities: *opened, turned, twisted, bend, passing*
Equipment	• Modifiers: ***[access]*** *ladder*	• Not directly used for construction: *lights* • Parts of equipment: *rung [of ladder]*
Physical_environment	• Material: ***[oak]*** *staircase* • Specificity: ***[foot of the]*** *stairs*	• General parts of sites/structures: *ground, site, wall*
Hazard	• Preposition: *slip* ***[off]*** • Verb object: *caught* ***[his foot]***	• Confusion *(Harmful_Consequence): fractured* • Confusion *(Construction_Activity)*: *bending down*

Table 3: Summary of recurring annotator discrepancies. Disputed span parts are **[emboldened in square brackets]**

Category	Total count
Body_part_injured	482
Harmful_consequence	194
Construction_activity	618
Equipment	980
Physical_environment	951
Hazard	608

Table 4: Annotation counts by category

ard concern confusion with other categories. Both *Hazard* and *Harmful_Consequence* are unexpected and undesirable occurrences, and as such, they include some degree of semantic overlap. This may explain why, in sentences where only an injury is mentioned, there was sometimes uncertainty about which category to choose e.g., *He* ***fractured*** *his vertebrae bone on his back*. Confusion between *Hazard* and *Construction_Activity* could occur when the activity directly leads to an injury, e.g., *The Technician involved dropped a spanner and whilst* ***bending down*** *to pick it up experienced some discomfort in his back*. Here, *bending down* is an everyday activity; even though it had consequences, it should not be annotated as a *Hazard*, as it was not an unexpected event.

5.3. Annotation Analysis

In Table 4, we show the total counts of each annotated category in the initial corpus. These statistics provide evidence that RIDDORs are very rich in descriptions of the contextual factors of incidents (i.e., the physical environment in which they take place and the equipment being used at the time). Since these contextual factors generally correspond to project attributes that are identified at the planning stage, there is scope for NER models trained according to our scheme to automatically identify RIDDORs that have most relevance to identifying potential risks in a new project.

Mentions of hazards are also fairly numerous (which is to be expected, given the nature of these reports), as are activities being undertaken at the time of the incident. This provides opportunities to explore how interactions between the physical aspects of a construction site and the actions undertaken on-site can result in the occurrence of incidents. The type of body part injured is mentioned for most hazards, although the number of harmful consequences is much lower, possibly because annotators were instructed not to annotate general words/phrases such as *injury*, since they do not provide a precise characterisation of the nature of the harmful consequence.

In Table 5, we show the most commonly marked up words and phrases in each category. We note that for some categories, the construction site attributes defined in Desvignes (2014) and Prades (2014) cover many of the phrases annotated. For example, the majority of the most commonly annotated *Construction_Activity* phrases in our corpus fall under the construction site attributes of *Drilling, Dismantling, Transitioning* and *Lifting/Pulling/Manual Handling*, while many of the most commonly annotated *Equipment* entities fall under attributes like *Light Vehicles, Scaffold, Stairs* and *Powered Hand Tool*.

Some of these construction site attributes correspond to fairly high-level concepts, and hence there is not necessarily a direct link between the phrase annotated in text and the related attribute (e.g., between the annotated entity *drill* and the attribute *Powered Hand Tool*). However, we note that these higher-level attributes could provide a useful means to group textual entities according to the type of concept that they describe. Table 5 provides evidence that even in this initial corpus, the annotations can provide evidence of how these attributes are realised in text in different ways; applying machine learning algorithms to detect additional entities with similar characteristics or which occur in similar contexts will identify further synonyms and/or terms related to these attributes, which are likely to go beyond the information covered in the attribute-specific, manually curated dictionaries of Tixier et al. (2016a). The application of an automatic normalisation method, e.g., Thompson & Ananiadou (2018), could help to map or *normalise* previously unseen entities to existing attributes, by taking into account their surface and semantic level similarities to existing dictionary entries.

While our annotation may thus be helpful in enhancing the recognition of mentions of this fixed set of construction site attributes, Table 5 also provides evidence that the annotation will be helpful in recognising a broader range of concepts. For example, while some of the *Physical_Environment* annotations are also covered by construction site attributes from Desvignes (2014) and Prades (2014), e.g., *Lumber[timber], Piping* and *Steel*, annotations

Body_part_injured	Harmful_conseq	Construction_act	Equipment	Physical_environ	Hazard
back	fracture	lifting	ladder	site	fell
hand	cut	drilling	vehicle	wall	slipped
left hand	broken	dismantling	scaffolding	floor	struck
wrist	twisted	removing	scaffold	ground	fall
shoulder	fractured	demolition	van	roof	hit
leg	dislocated	moving	step ladder	joist	caught
arm	laceration	lowered	drill	timber	tripped
ankle	cutting	lift	platform	beam	lost balance
finger	bruised	walking	fixed scaffold	first floor	fell over
thumb	break	removed	stepladders	steel	falling
head	broke	unloading	steps	slab	stumbled
ribs	breaking	cutting	tower scaffold	boards	slip
right hand	jarred	dismantle	vehicles	pier	landing
knee	fracturing	fitting	blade	pipe	trapped

Table 5: Most commonly annotated spans in each category

of more specific parts of structures and/or locations are generally not covered by these attributes, nor are some *Construction_Activity* terms, such as *fitting* and *cutting*.
Since the attributes of Desvignes (2014) and Prades (2014) are mostly restricted to features of the construction site (i.e., activities, materials and equipment), they do not generally cover concepts that we identify in the categories of *Body_part_injured, Harmful_consequence, Hazard* and *Protection_Measure*. Hence, our scheme and annotations can help to collect evidence about how these concepts are mentioned in domain-specific documents, which is particularly relevant, given their demonstrated importance in modelling domain-specific knowledge.

6. Conclusion and Future Work

Motivated by the current scarcity of semantically annotated corpora for the construction safety domain, we have described the design of a novel NE annotation scheme that is specifically aimed at identifying mentions of concepts that are highly pertinent to identifying hazards, their contextual factors, their consequences and possible means to mitigate them. Corpora annotated according to the scheme are intended to facilitate the development of tools to support a range of practical tasks, such as semi-automatic generation of risk registers and integration of structured health and safety information into BIM software. An initial effort to annotate 600 sentences from RIDDOR accident reports allowed us to verify the utility and suitability of the scheme in two ways. Firstly, we showed through an average IAA of 0.79 F-score that the scheme can be applied by different annotators in a fairly consistent manner; our study of annotator discrepancies will allow us to further refine the guidelines. We can thus aim for an even greater degree of consistency when we carry out planned work to produce a much larger corpus annotated according to the scheme, which will include a wider range of documents from multiple relevant sources. Secondly, we found that most categories of entities in the scheme occur with reasonably high regularity in text, and can thus provide evidence of how a wide range of relevant concepts can manifest themselves in text. Especially in conjunction with a larger corpus, the application of machine learning techniques has the potential to push the potential range of concepts that can be discovered automatically considerably beyond the possibilities of existing tools.
Once we have completed annotation of a larger corpus, we aim to pursue two additional lines of work. The first of these will be to normalise entity mentions, in order to map them to the exact types of concepts that they represent. Although, as mentioned above, it may be useful to map certain entities to previously-proposed construction site attributes, the complete range of concepts that we cover extends beyond these attributes. Thus, we will need to consider other ontologies and/or terminological resources, which may include: the Industry Foundation Classes (IFC) (ISO, 2013), which covers building elements at various levels of granularity, and is used in BIM; Uniclass (NBS, 2015), which includes various types of activities, products, tools and locations; and clinical/medical ontologies, such as SNOMED CT (Donnelly, 2006), which covers body parts and injuries. Secondly, we will create a knowledge graph by extracting and structuring knowledge contained in construction safety documents. An end-to-end model (Mesquita et al., 2019) will be implemented to extract NEs and relations between them. We will enrich the domain-specific entities detected with models trained using our corpus with the aid of resources such as ConceptNet (Speer et al., 2017) and DBpedia (Lehmann et al., 2015), which, despite being concerned with general language, may still help to detect entities that are relevant within this domain. Subsequently, we will follow an Open Information Extraction approach to the discovery of relations (Cetto et al., 2018), which can extract a wide range of relation types, without requiring a training corpus in which relations have been manually annotated.

7. Acknowledgements

We thank Belinda Oakley, Carl Gartside, Claire Collins and Claudia Tarr of HSE for their dedicated annotation work.

8. Bibliographical References

Al Qady, M. and Kandil, A. (2009). Concept relation extraction from construction documents using natural language processing. *Journal of Construction Engineering and Management*, 136(3):294–302.

Al Qady, M. and Kandil, A. (2014). Automatic clustering of construction project documents based on textual similarity. *Automation in Construction*, 42:36–49.

Bertke, S., Meyers, A., Wurzelbacher, S., Bell, J., Lampl, M., and Robins, D. (2012). Development and evaluation of a naïve bayesian model for coding causation of workers' compensation claims. *Journal of Safety Research*, 43(5-6):327–332.

BSI. (2018). PAS 1192-6: 2018: Specification for collaborative sharing and use of structured health and safety information using BIM.

Caldas, C. H. and Soibelman, L. (2003). Automating hierarchical document classification for construction management information systems. *Automation in Construction*, 12(4):395–406.

Cetto, M., Niklaus, C., Freitas, A., and Handschuh, S. (2018). Graphene: Semantically-linked propositions in open information extraction. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2300–2311.

Chen, H. and Luo, X. (2019). An automatic literature knowledge graph and reasoning network modeling framework based on ontology and natural language processing. *Advanced Engineering Informatics*, 42:100959.

Chi, N.-W., Lin, K.-Y., and Hsieh, S.-H. (2014). Using ontology-based text classification to assist job hazard analysis. *Advanced Engineering Informatics*, 28(4):381–394.

Desvignes, M. (2014). Requisite empirical risk data for integration of safety with advanced technologies and intelligent systems. Master's thesis, University of Colorado at Boulder.

Donnelly, K. (2006). SNOMED-CT: The advanced terminology and coding system for eHealth. *Studies in health technology and informatics*, 121:279–290.

Esmaeili, B. and Hallowell, M. (2012). Attribute-based risk model for measuring safety risk of struck-by accidents. In *Proceeedings of the Construction Research Congress 2012: Construction Challenges in a Flat World*, pages 289–298.

Fan, H. and Li, H. (2013). Retrieving similar cases for alternative dispute resolution in construction accidents using text mining techniques. *Automation in construction*, 34:85–91.

Gao, G., Liu, Y.-S., Lin, P., Wang, M., Gu, M., and Yong, J.-H. (2017). BIMTag: Concept-based automatic semantic annotation of online BIM product resources. *Advanced Engineering Informatics*, 31:48–61.

Glazner, J., Bondy, J., Lezotte, D. C., Lipscomb, H., and Guarini, K. (2005). Factors contributing to construction injury at Denver International Airport. *American Journal of Industrial Medicine*, 47(1):27–36.

Goh, Y. M. and Ubeynarayana, C. (2017). Construction accident narrative classification: An evaluation of text mining techniques. *Accident Analysis & Prevention*, 108:122–130.

Guo, B. H. and Goh, Y. M. (2017). Ontology for design of active fall protection systems. *Automation in Construction*, 82:138–153.

Hripcsak, G. and Rothschild, A. S. (2005). Agreement, the F-Measure, and reliability in information retrieval. *Journal of the American Medical Informatics Association*, 12(3):296–298.

ISO. (2013). Industry foundation classes (IFC) for data sharing in the construction and facility management industries.

Ju, M., Short, A. D., Thompson, P., Bakerly, N. D., Gkoutos, G. V., Tsaprouni, L., and Ananiadou, S. (2019). Annotating and detecting phenotypic information for chronic obstructive pulmonary disease. *JAMIA Open*, 2(2):261—-271.

Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., Van Kleef, P., Auer, S., et al. (2015). DBpedia–a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 6(2):167–195.

Li, M., Nguyen, N., and Ananiadou, S. (2017). Proactive learning for named entity recognition. In *Proceedings of BioNLP 2017*, pages 117–125.

Mesquita, F., Cannaviccio, M., Schmidek, J., Mirza, P., and Barbosa, D. (2019). KnowledgeNet: A benchmark dataset for knowledge base population. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 749–758.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Miller, G. (1995). WordNet: A lexical database for english vol. 38. *Communications of the ACM*, 38(11):39–41.

NBS. (2015). Uniclass 2015–a universal classification system for the construction industry.

Nghiem, M.-Q. and Ananiadou, S. (2018). APLenty: annotation tool for creating high-quality datasets using active and proactive learning. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 108–113.

Prades, M. (2014). Attribute-based risk model for assessing risk to industrial construction tasks. Master's thesis, University of Colorado at Boulder.

Settles, B. (2009). Active learning literature survey. Technical report, University of Wisconsin-Madison Department of Computer Sciences.

Speer, R., Chin, J., and Havasi, C. (2017). ConceptNet 5.5: An open multilingual graph of general knowledge. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 4444–4451.

Taylor, J. A., Lacovara, A. V., Smith, G. S., Pandian, R., and Lehto, M. (2014). Near-miss narratives from the fire service: a Bayesian analysis. *Accident Analysis & Prevention*, 62:119–129.

Thompson, P. and Ananiadou, S. (2018). HYPHEN: a flexible, hybrid method to map phenotype concept mentions to terminological resources. *Terminology*, 24(1):91–121.

Thompson, P., Daikou, S., Ueno, K., Batista-Navarro, R., Tsujii, J., and Ananiadou, S. (2018). Annotation and detection of drug effects in text for pharmacovigilance. *Journal of Cheminformatics*, 10(1):37.

Tixier, A. J.-P., Hallowell, M. R., Rajagopalan, B., and Bowman, D. (2016a). Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports. *Automation in Construction*, 62:45–56.

Tixier, A. J.-P., Vazirgiannis, M., and Hallowell, M. R. (2016b). Word embeddings for the construction domain. *arXiv preprint arXiv:1610.09333*.

Tixier, A. J.-P., Hallowell, M. R., Rajagopalan, B., and Bowman, D. (2017). Construction safety clash detection: identifying safety incompatibilities among fundamental attributes using data mining. *Automation in Construction*, 74:39–54.

U.S. Department of Labor Bureau of Labor Statistics. (1997). Occupational injury and illness classification manual.

Zhang, J. and El-Gohary, N. M. (2013). Semantic nlp-based information extraction from construction regulatory documents for automated compliance checking. *Journal of Computing in Civil Engineering*, 30(2):04015014.

Zhang, S., Boukamp, F., and Teizer, J. (2015). Ontology-based semantic modeling of construction safety knowledge: Towards automated safety planning for job hazard analysis (JHA). *Automation in Construction*, 52:29–41.

Zhang, F., Fleyeh, H., Wang, X., and Lu, M. (2019). Construction site accident analysis using text mining and natural language processing techniques. *Automation in Construction*, 99:238–248.

Zou, Y., Kiviniemi, A., and Jones, S. W. (2017). Retrieving similar cases for construction project risk management using natural language processing techniques. *Automation in Construction*, 80:66–76.