

過去のヒヤリハット事例のデータ解析に基づく 再発防止のための有用情報提供に関する研究

Research on providing useful information to prevent recurrence
based on data analysis of past near-misses

上野 樹生

指導教員 鈴木 雅人, 北越 大輔, 西村 亮

東京工業高等専門学校 情報工学科 知識情報研究室

キーワード：類似キーワード推定, TF-IDF, ヒヤリハット, 安全管理

1. はじめに

共同研究を行っているビル内空調設備施工会社（以下「協力企業」）では、電気・空調類の管理や施工業務を行っている。施工現場には危険な要素があふれている（図1）ため、“ヒヤリハット”が発生する要因は多く、それらを未然に防ぐためにも、十分な管理体制を整えることが大事である。しかし、現在利用しているヒヤリハット報告書は、発生日・発生事例・解決策の3項目のみを記述したものが紙媒体で保存してあるだけなので、他の従業員が確認をしたり、過去の事例報告書を活用したりすることは殆ど行われていないのが現状である。

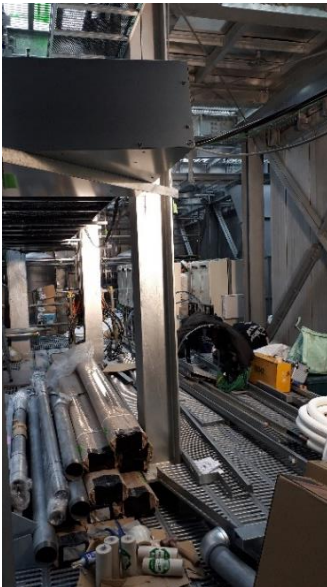


図1 実際の施工現場

そこで本研究では、過去の事例をもとに、これから施行予定の現場状況や企業特性に合わせた、過去のヒヤリハット事例の提供をする重大事故未然防止システムを構成する。そして、利用者へ安全管理上のアドバイスを与えることで、ヒヤリハット

の再発防止を実現する。

2. 重大事故未然防止システム

本システムでは、協力企業から提供された約2,000件のヒヤリハットサンプルデータ（以下「サンプルデータ」）を解析してデータベースとして管理し、それを用いて、過去の事例を提供して注意喚起するシステムを提案する。

本システムは、図2のようにサンプルデータからキーワードを抽出し、TF-IDF値[1]を求める部分と、図3のようにデータベースから適切なヒヤリハット事例を選択し提示する部分の2段階構成となっている。以下の節より各段階の詳細を示す。

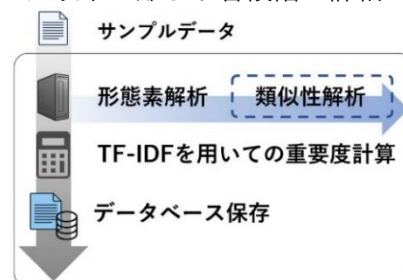


図2 キーワード抽出からTF-IDF値計算の流れ



図3 提示事例選択の流れ

2.1 キーワードの抽出と TF-IDF 値の計算

サンプルデータの「発生事例」文章から、内容を解析するうえで必要なキーワードを選定する。形態素解析を用いて、発生事例の文章を単語レベルに分割した上で、「はしご」や「工具」などの出現頻度が高い名詞や、「転ぶ」や「走る」などの出現頻度が高い動詞を、発生事例ごとにキーワードとしてデータベースに追加する(表1)。また、「～時」などのタイミングを表す単語もデータベースに追加する。その際、類似性のある単語を調査し、それらは同じ単語として扱う。

表1 事例文章からのキーワード抽出イメージ図

1	サビステーションに足を踏み入れた時、廻りが暗く段差に気がつかずバランスを崩して転びそうになった。	サビステーション/足/踏み入れ/暗く/段差/バランス/転びそう
2	アップスターの梯子に昇っている時、吊り下げてあった全ネジに気付かずそのまま上階まで直撃	アップスター/梯子/昇って/吊り下げ/ネジ/気付かず/脳天/直撃

次に、キーワードの TF-IDF 値を計算する。この値により、そのキーワードに関するヒヤリハットが頻繁に起きているかどうかを解析できるだけでなく、そのキーワードが出現する事例が、全事例内でどのくらい希少性が高いかを解析することにより、普遍的に出現しない単語の重要性も考慮することができる。尚、各事例での TF-IDF 値はキーワードに対応させたベクトル形式で、データベースに保存する。

2.2 TF-IDF 法を用いた提示事例の選択

利用者は、過去のヒヤリハット事例を得るために、0もしくは1個の基本タグを選択し、必要に応じて1個以上の自由記述キーワードを追加する。そして事例文章ごとに、入力されたキーワードの TF-IDF 値により表示スコアを決定する。またこの際に、基本タグを考慮したスコアを算出する。

(1) 基本タグ

協力会社の業務の内容に鑑み、出現頻度の高いキーワードのうち、「運搬時」「施工時(空調)」「施工時(配管)」「配線時」の4種類を基本タグとする。これらの基本タグを含む事例の文章のスコアを1

追加することで、より提示されやすくする。

(2) 自由記述キーワード

利用者が自由記述キーワードを入力した場合は、入力キーワードとデータベースのキーワードとの類似度を計算したうえで、最も類似しているキーワードを選定し、その TF-IDF 値を用いてスコアの計算を行う。

以上の(2)自由記述キーワードでの TF-IDF スコアと、選択されている(1)基本タグにより候補スコアに追加される+1を足し合わせて事例文章ごとにスコアを計算し、スコアが降順になるようにソートして発生事例の提示を行う。このことにより、分かりやすいインターフェースで、過去事例を紹介し、利用者への注意喚起を行うことができる。

3. まとめ

本研究では、協力企業の特徴に合わせたヒヤリハット再発防止の達成のために、類似キーワードの推定並びに TF-IDF 法を活用して、利用者に分かりやすく事例紹介するシステムを構築した。また、利用者に、できる限りの安全に対する意識改革を行うため、本システムは「簡単に使用できる」ことや「継続的に使用できる」ことを念頭に、出力データを「分かりやすく」表示することに重点を置いて開発した。今後は実際の現場で試用することによりシステムの評価、改善を行う予定である。

謝辞

本研究を進めるにあたり、サンプルデータの提供をして頂いた、協力会社の関係者の方々に深謝の意を表します。

参考文献

- [1] 河内 沙織, 豊田 哲也, 延原 肇. “Wikipedia カテゴリの関連度と要点情報の複合評価指標に基づく論文キーワード抽出”. 27th Fuzzy System Symposium. 2011, vol 27, TE1-1, p. 671-675.