Automating Incident and Accident Analysis

Mr. Fergus Toolan
Intelligent Information Retrieval Group
University College Dublin
Ireland

2nd Bieleschweig Workshop on Root Causal Analysis

Braunschweig, Germany – July 1st-2nd, 2003

Overview

- Motivations for automation
- Incidents / Accidents in terms of automation
- Methods of automation
- Data Mining
- Automated Root Causal Analysis
- Evaluation

Motivations

- Why do we wish to automate?
  - Less person hours involved
  - Analyse more incidents
  - Use automation as a tool to aid the experts
  - Benefit to safety as more incidents are studied in greater depth.

Incidents / Accidents

- All familiar with the terms
- The distinction between incidents and accidents lead to different types of automation
- Incidents can be viewed as frequently occurring events which are logged but rarely investigated.
- Incidents contain more field based than textual information e.g. ASRS Dataset

Incidents / Accidents

- Accidents are more serious than incidents
- They result in more investigation / analysis
- Hence they provide more long written reports than incidents provide
- This leads to a different analysis technique for both domains

Methods of Automation

- With the differing natures of the two types of incident identified we have many different methods that are applicable to the analysis

Incidents:	Accidents:
Case-Based Reasoning (CBR)	Textual CBR
Data Mining	Text Mining
Classification	Information Retrieval
	Information Extraction

Data Mining

- Not going to go into specifics.
- Basically Data Mining is a Machine Learning Technique that allows the discovery of patterns in large datasets.

Automated Root Causal Analysis

- Automated Root Causal Analysis (ARCA) is a tool to aid the user in performing a root causal analysis on an accident.
- It can be divided into three major phases
  - Event / State Identification
  - Temporal Identification
  - Causality Discovery

Event / State Identifcation

- Method
  - Remove Stopwords, Perform Stemming
  - Divide Report into sections
  - Cluster Sections based on Cosine Similarity
  - Extract the most frequently occurring n-grams from these clusters
  - These are candidates for events / states

Temporal Identification

- Method
  - NLP techniques applied to original report to try and order the events / states identified.
  - Use of words such as before, after, concurrently,… (temporal words)
  - Information Extraction approach – accident reports are semi-structured so it may be possible to get an idea of sequence through the report layout

Causality Discovery

- Very difficult phase;
- Possibilities:
  - We can identify possibilities for causality.
  - For instance if A occurs before B it tells us
    - B could not have caused A
    - A ***might*** be a cause of B

Preliminery Evaluation

- Event / State Identification
  - Ladbroke Grove
- Data Mining
  - IAAU Dataset – 69 incidents
  - ASRS Dataset – 2050 incidents

Questions???

Comments!!!