Design Guidelines for Exploring Relationships in a Connected Big Data Environment

Jaison Jacob, Santhosh Rao

Design guidelines for exploring relationships in a connected big data environment

Jaison Jacob[1] and Santhosh Rao[1]

SAP Labs India[1], 138 EPIP, Whitefield Bangalore 560066, India
{jaison.jacob01, santhosh.rao}@sap.com

Abstract. Reimagining the 'SAP Investigative Case Management' frame-work from a log-based register of events to a direct interaction environment with the possibility to search, explore relationships between multiple enti-ties in one or more cases/incidents. This case study is about our approach in conceptualizing a generic network visualization method by deconstructing the existing data models. We devised a set of guidelines that can be employed to represent a large number of entities with the intention of examining their relationships.

Keywords: Big-data. Network visualization framework. Visualizing connected data.

1 Introduction

SAP Investigative Case Management is a SAP CRM[1] solution used by an investigator, detective, or user from law enforcement agencies to report a crime or offence or to investigate or probe a crime or offence. There are two types of incident reporting tools, viz. (1) systems for cases where there is a dispute between two people (2) systems for cases that require exploration of a network that has over 5000 entities. This type of data is dynamic in nature and grows over time. All of it can be thought of as "big data". Our secondary research [1] says that there are 23 types of incident reporter tools that an investigator can use to report an incident, and there exists a common task pattern amongst them: (1) **Users detect an incident.** On the event of an incident, the user travels to the incident location to gather artefacts. (2) **They create an incident report.** The incident report contains a description, location, date and time of incident occurrence, and user identification. (3) **Users follow up the incident to share information with other users.** Depending on

[1] SAP CRM is an integrated customer relationship management (CRM) software manufactured by SAP SE that targets business software requirements of midsize and large organizations in all industries and sectors.

the condition of the report, there may be organizational restrictions on sharing the outcome of the report once it is registered.

1.1 State of art

There are several disadvantages in existing systems[2]: (1) **Limited parameters to describe/report an incident.** More parameters or entities = more detailed data. (2) **Report keeping and lack of categorization.** Storing a summary of incidents with no possibility to cross reference is of little or no use. (3) **Not configurable based on context.** If the reporter wants to add a new field called "sexual assault" with "men" as victims w.r.t the case details, it's not possible, because according to the existing system, sexual assault victims can only be female.

SAP Investigative Case Management or ICM improved this existing system by providing a configurable entity model with more entity 'fields', to empower the reporter to enter detailed data about the incident. It has a search feature that revolutionizes "incident probing". Now the case worker can go through historical records, based on very specific entity type search. With this the system moved from a 'incident reporting system' to a 'probing system'

The improved list of entity types is as follows: case, lead, geo-location, objects associated with the incident, person and organizations, incidents and activity. Above all, ICM focused on relationships between cases. A reporter can now create associations between data entities using relationships and then rate the reliability of the linked data using a reliability matrix. Here the reliability of the information and source of the information are evaluated. This information can pertain to a relationship, or to a description, or to the profile of a person. For e.g., if a relationship exists between a suspect and a victim based on a witness account, you would be able to set a reliability level for the witness, and another for the relationship between suspect and victim.

[2] **Existing incident reporting systems**. Incident information is reported, more or less, by employing the following entities: (1) Incident Description. Textual description of the incident or by choosing a description from a list of predefined incident types. In some cases, description is accompanied by a picture of the incident. (2) Time of incident occurrence. (3) Incident location. This includes a "geo location pin", where the user can pinpoint the incident or an address of the location supported by landmarks. [1]

2 Approach

But ICM had their own problems. It was difficult to define a new relationship in the system as the user had to search and find cases that are similar in nature, go through them and deduce who is related to what and then define relationships based on this knowledge.

Reduce the ICM framework to 'atoms'. The core principle is to connect an incident with people; both are always related by a time stamp and a geolocation. The incident can later be drilled further into an object and a location associated with an incident. To detail the above classification, we came up with a 'non-hierarchical' classification of all the entities, (a "node"). A 'node' can be a person, a location, an event or an object involved in the incident.

Deconstruct the ICM framework by creating a network of atoms. A 'node' can have a 'relationship' with 'another node'. It is a representation of how different types of entities are related to each other from the perspective of an incident.

The network visualization framework. A network diagram is a set of entities exhibiting linear as well as non-linear relationships, graphically represented as nodes (entities) connected with lines (relationships). A network of nodes can exhibit a non-hierarchical distribution, to help the user simplify a complex relationship network and vice versa. A node can be related to one or more nodes. A node can contain sub-nodes and these sub-nodes can contain more nodes and so on.

3 Design Guidelines

Tell a story using data. Every selected node in the UI should answer three questions: 'When'(time), 'Where'(geolocation) and 'How' (relation with other nodes in network). To perform this task, the user should have the right set of actions at the right time - "contextual menu". The menu should accommodate the following actions: expand nodes, collapse nodes, show information about the node in focus, select multiple nodes (for comparison), delete (non-related) nodes.

Power of choice and importance of probing path. Often, multiple nodes surround the node in focus. Now the user has to make a decision on which node to select to explore further. Each choice made, can lead to more nodes and eventually builds a path. This path can tell how the investigation progressed over a period of time.

Direction of the probing path matters. The network exploration path is important to understand how nodes are related. Relationships change with user's perspective.
Flexibility to zoom-in and out. UI should scale to accommodate details of a growing network. The user should be able to simplify the view so that he can see only the necessary information: Natural Zoom: Zoom in and Zoom out, Semantic Zoom: When a user zooms out of a canvas with a number of selected nodes, only these selected nodes and the relationships should be highlighted. The rest can be hidden. Clustering for nodes in geolocation: when user zooms out in a map view, the nodes on the map should exhibit clustering. All similar nodes should be replaced with a single representative node.
Divide and rule across data layers. Allow the user to choose visual filters: people, object, location and events. Possibility to view nodes on the canvas in an appropriate environment. Based on the context, show the nodes on a map, or as a network, or on a timeline.
Non-biased UI and the proof for taking 'informed' decision. The system should capture the probing path since big data is dynamic in nature. Since it's an informed decision, the decision changes with the UI. The system should not bias the user while making his decisions. It should only aid him in deducing results.

4 Conclusions and future work

The network visualization framework can be applied on any set of data that exhibits linear as well as nonlinear relationships. E.g., Healthcare domain. A doctor can view the patient's medical history by way of a network, use a timeline to organize a series of medical incidents. The doctor can see how these incidents are related with each other with respect to time, the doctors involved, diagnosis and medical reports. Similarly, we can also employ such a network in a scenario to help a recruiter look for candidates to fill a specific role.

5 References

1. Cedric Bach., Regina Bernhaupt., Caio Stein D'Agostini., Marco Winckler of Mobile applications for incident reporting systems in urban contexts: lessons learned from an empirical study, ECCE'13 Proceedings of the 31st European Conference on Cognitive Ergonomics, Article No. 29, in ACM, (2013)