Relationship To Other Research Support

My research area is empirical software engineering and the goal of my ongoing research program is to understand how and help developers to produce successful software systems. I am also interested in statistical machine translations and summaries of code identifiers using freeform text, such as StackOverflow documentation.

The main goal of my NSERC Discovery grant, Contemporary Software Peer Review: Modern practices, fault prediction, and extraction of design decisions is to understand the software review practices in industrial firms as opposed to open source projects I studied in my thesis (from this grant, I have published a paper at the top tier conference Foundations of Software Engineering). In the proposed Engage grant, I will be able to use my expertise and scripts to extract Ericsson data. A possible synergy between the grants, would be to use peer review data in models for test prioritization.

My Department of National Defense (DND), NSERC, industry (KDM Analytics) grant, The Impact of Disruptive Events on Software Systems, involves studying disruptive events that lead to poor software outcomes. A disruptive event, such as developer turnover, will have risks and mitigating factors that we are measuring. While there is no direct overlap between the grants, it may be possible to observe how test failures disrupt the software development and release processes.

The goal of my FRQNT grant is to translate code elements, classes and methods, from English into French. This grant does not overlap with the Engage proposal. The identifier names (eg names of classes and methods) of most major software libraries are based on English terms that capture the purpose of identifiers (eg the Android 'AccountManager' class). There are thousands of identifiers on the Android project, putting non-English speakers at a disadvantage. Statistical translations (eg Google Translate) of technical documents using non-technical language models, results in incorrect translations of technical terms (eg the term 'Window' has a technical meaning quite different from its non-technical meaning). Unlike non-technical document translation, we have observed that library identifiers are not translated when used in multiple languages, 'AccountManager.addAccountExplicitly qu'est-ce que cette fonction fait?' I am in a unique position to create language models to translate software documents because, as a postdoctoral research, I developed a technique that extracts identifier names in freeform text and code fragments. We can take two comparable corpora, such as, the community forums on Android in French and English and statistically determine which words tend to co-occur with each identifier. These co-occurring words represent a language model that describes each identifier in its respective language, which when aligned, allows us to 1) describe the purpose and behaviour of an identifier 2) increase the quality of translations of entire documents that discuss identifiers.