***Collaboration Plan***

***Background and Expertise***

The proposed work involves close collaboration between Tim Menzies and Xipeng Shen and our industrial colleagues at Microsoft, IBM, and Lexis Nexis.

The PIs have complementary expertise.

* Tim Menzies has extensive experiences in Software Engineering (SE) and data analytics: since 2002, he has been mining software project data for large organizations such as NASA and Microsoft. His recent work includes extending analytics to multi-objective optimization as well as editing several prominent textbooks documenting the international software analytics community.
* Xipeng Shen has extensively worked on program compilation, software autotuning, program compile-time and run-time optimizations. His recent work covers the designs of specification API and optimizing compilers for data analytics applications.

Our industrial colleagues at Microsoft, IBM, and LexisNexis bring much needed industrial perspective to this process:

* Throughout the following, we will conduct monthly teleconferences with our industrial colleagues who conduct software analytics studies at Microsoft, IBM and LexisNexis. Using those teleconferences, we will
  + 1. collect the work flow knowledge needed for our research;
    2. debate what software analytics toolkit we should be used for exploiting synergies.
    3. decide what software analytics methods deserve most optimization;
    4. decide what SE data sets should be analyzed by those methods.

***Relationship of PI’s Backgrounds to the Project***

The proposed project requires expertise in SE, data analytics, program compilation, code optimizations and runtime adaptation, and software autotuning. The two PIs complement each other and jointly cover all of these areas – with Menzies’ background in SE and data analytics, and Shen covering the other areas.

***Collaboration Plan***

This project involves two PIs in the same institution. The distribution of work between the two sites will be as follows. PI Menzies and his two graduate students will focus on the expansion of WHERE to support a broad range of tasks and to leverage WHERE for fast search in tuning. PI Shen and his two graduate students will focus on the design and implementation of SEAS, and the implementation of the grey-box tuner for exploring the synergy across levels of options in the tuning process. All the PIs and students will collaborate in the integration of the techniques into a single unified toolkit for holistic scalable autotuning.

The two PIs work in the same department and have had research collaborations before (a co-authored paper on SE autotuning at [Information & Software Technology 76](http://dblp.uni-trier.de/db/journals/infsof/infsof76.html#FuMS16)). For this research project, they will conduct weekly meetings. All personnel (PIs and graduate students) involved in the project will participate.

The entire team will set up a code repository using a standard versioning software (e.g. GIT) to manage code development and keep track of software changes. Results will be tracked via the software repository, journal papers, and conference presentations.

***Proposed Timeline***

The proposed project will be carried out over four years, with the following time-line.

* **Year 1:** (a) Start the monthly teleconference series with our industrial colleagues; (b) Designing and implementing SEAS based on distance-related analytics problems; (c) finishing the basic framework of the synergies queries and equipping it with some basic algorithmic options; (d) extending GALE to support tabu search and forms the base of our learner SMOBA; (e) exploring the usage of SMOBA for characterizing input datasets of SE data analytic applications.
* **Year2:** (a) Enriching SEAS with a broader range of algorithmic choices for SE data analytics (beyond distance-related algorithms); (b) developing APIs and compiler support for the grey-box tuner and exploiting the synergy between the compilation-level options and algorithmic choices; (c) expanding SMOBA to support response-surface methods and bilevel optimizations; (d) investigating the usage of SMOBA for locality-based fast search in autotuning.
* **Year 3:** (a) Incorporating into SEAS and the grey-box tuner the optimizations of the options at the execution level and investigating their synergy with the compile- and algorithm-level tunings; (b) further developing SMOBA to support anomaly detection and incremental adjustment; (c) leveraging SMOBA to facilitate efficient autotuning by finding and avoiding redundant searches;
* **Year 4:** Refactoring all the code over all the years work to create a full version 1.0 of our toolkit, documenting it w.r.t known and interesting SE data analytics cases and real-world industrial cases.