Google Your Life: Learning Sensors Data
Dan Feldman, Massachusetts Institute of Technology

Your smart-phone can collect real-time data about your life using its sensors. This includes where you are, what you see, and what you hear in any given moment. Tera-bytes of such data are written to smart-phones all over the world in every given second.

We are developing algorithms and systems to turn the sensors data into information in the form of a readable diary. The diary enables the automatic creation of an auto-biography for the user. Additionally, the system provides real-time summaries of daily activities (e.g., which restaurants you go to and when, how you spend your leisure and work time, etc.) that can be shared with friends and family.

How can we turn noisy sensors signals into a searchable text? How to learn our users statistics while preserving their privacy? We answer these questions by maintaining a semantic compression of the streaming data (called sketch or core-set). This core-set represents the original data in the sense that running queries or fitting models on the semantic compression will yield a similar result when applied to the original data set, under natural assumptions (intuitively, that you are not doing a random walk on the planet).

Combining map-and-reduce techniques with our core-sets yields a system capable of compressing in parallel a stream of $O(n)$ samples using space and update time that is only $O(\log n)$. We present the current state of our application, experiments and theoretic results.