Incremental maintenance of queries on dynamic trees

Antoine Amarilli, Charles Paperman

The task of *incremental maintenance on dynamic data* studies how we can efficiently evaluate queries and then quickly update their results whenever the data is modified. One simple example is queries that ask whether the data belongs to a specific formal language L. For instance, given a word w, we can test in time $O(|w|)$ whether $w \in L$: then w is modified, e.g., by changing characters, and we want to know whether $w \in L$ after each modification. The naive algorithm is to test membership to L by reading the whole word after each update, with update complexity $O(|w|)$: but in many cases we can do better. For instance, consider the language $b^*(ab^*ab^*)^*$ of words on $\{a, b\}$ with an even number of a's. We can maintain membership to this language in $O(1)$ — simply by keeping a count of the number of a's.

We have recently studied this incremental maintenance problem for *regular languages* on *words* [AJP21], where we were able to precisely pinpoint the update complexity for different languages. We showed that membership to regular languages can always be performed in time $O(\log |w| / \log \log |w|)$ per update, and that this bound is tight for some languages. We identified a class of languages with complexity $O(1)$ per update (e.g., finite languages), and identified intermediate languages with update complexity $O(\log \log |w|)$ and with a conditional superconstant lower bound. These results are obtained by combining methods from algebraic automata theory with algorithmic lower bound techniques.

The goal of this internship is to study the extension of these results to *regular tree languages*, defined via tree automata. These languages can be studied via the formalism of *forest algebras*, and also by leveraging some new algorithmic lower bounds. For instance, results on the *marked ancestor problem* imply a $\Omega(\log n / \log \log n)$ bound for some languages [AHR98], where n is the size of the tree. We have achieved first results in the area, but many open directions remain. One goal could be to characterize the various possible complexity regimes under the assumption that tree languages admit a so-called *neutral letter*, and that the forest algebra is *aperiodic* in a certain sense.

Many other directions can be explored, such as:

- Studying concrete examples of tree languages, for which ad hoc algorithms or reduction techniques can be designed.

- Considering various kinds of updates: from relabeling updates to insertion/deletion updates, with possible connections to tree balancing schemes.
- Investigating incremental maintenance for *context-free languages*, where the tree structure is not given but must be obtained via parsing.

Supervision and environment. This internship is intended for a duration of 5 months, from February to June. It will take place in the LINKS team at Inria Lille in the north of France, and will be cosupervised by Antoine Amarilli[1] (Advanced Research Position at Inria Lille) and Charles Paperman[2] (Associate professor at Université de Lille).

Applications should be sent by email to the two supervisors: `a3nm@a3nm.net` and `charles.paperman@univ-lille.fr`.

References

[AHR98] Stephen Alstrup, Thore Husfeldt, and Theis Rauhe. Marked ancestor problems. In *FOCS*. IEEE, 1998.

[AJP21] Antoine Amarilli, Louis Jachiet, and Charles Paperman. Dynamic membership for regular languages. In *ICALP*, 2021. Best paper award of ICALP track B.

[1] `https://a3nm.net/`
[2] `https://paperman.name/`