Re: Mark W. Hopkins theory perspective on parser engine technology?

Source: http://coding.derkeiler.com/Archive/General/comp.theory/2008−04/msg00096.html

---

- *From*: Rock Brentwood <markwh04@xxxxxxxxx>
- *Date*: Fri, 25 Apr 2008 12:41:59 −0700 (PDT)

On Mar 28, 11:58 am, Chris F Clark <c...@xxxxxxxxxxxxxxxxxxxx> wrote:

> However, he also reaches some results that are contrary to established proofs. For example, if I recall correctly, he claimed at 1 time to have a linear parsing algorithm for a general CFG, which is well known to be O(n**2).

Section 1 of the "Algebraic Approach" is now, essentially, part of the Lecture Notes in Computer Science, 4988. A presentation for the material was given at RelMiCS 2008 earlier in April. An on−line version (with ad−supports mixed in by the server, sorry about that) is:

The Algebraic Approach I & II
http://federation.g3z.com/CompSci/RelMiCS/MarkH/Cover.htm

The conference itself plus some of what went on:

RelMiCS 2008
http://federation.g3z.com/CompSci/RelMiCS/index.htm

I'm going to add more detailed synopses and discussions of all the papers (and tutorials) presented there, hopefully soon.

Eventually, this site will be converted to an interactive blog when the ad−support is finally lifted out of it.

On the issue itself:

Recognition is O(n^log 7), not O(n^2) ... unless something new was just developed. Parsing −− as specifically defined in the references there −− is indeed O(n); as is the general problem of representing the set of translations for a given input. That comes directly out of Kolmogorov complexity theory and is, in fact, a simple (indeed, trivial) result.

For the recognition problem, the O(n^log 7) algorithm comes directly out of the O(n) output string described in those references (which is a context−free expression) by reduction to normal form over the pure Bra−Ket algebra. This reduction, in fact, directly incorporates the Boolean Matrix multiplication process that underlies modern variants of the Valiant algorithm (which tie to the shortest paths problem) and is closely connected to the shortest−paths formulation. In fact, you might consider it the algebraic reformulation of Valiant, itself.

The O(n) representation is farily clear and (almost) trivial. It's the same punchline you see in Kolmogorov theory. Once you see it you end up saying "oh, that".

There are other O(n) representations which are probably even more anti−climatic and more trivial. For instance, one can simply adopt the notation (T:w) for a translation grammar T and input word w to express the representation of the set of outputs translated from w. That's obviously O(n) in the length of the word, but less useful (for instance) for extracting the individual items out of the set (the "enumeration" problem) or for even determining the emptiness of the set, itself (the "recognition" problem).

In effect, that's been the approach adopted from the 1970's on, when writing down a calculus for context−free expressions. Nowadays, it's called the mu−calculus. You'll see, by the way, in section 3.7 of the "Algebraic Approach" article, the mu−calculus was combined with the bra−ket algebra, forming a highly fluid syntax that allows you to go back and forth between the two and mix mu expressions with bra−ket expressions.
.