Formal Language, Intuition, Total Order, Kleene plus and Zorn's lemma

Minseong Kim

In computer science, a character set Σ is often defined. Then, Kleene plus and Kleene star for formal language are defined. Then, $\Sigma^+ = \Sigma^*\Sigma$ is proved, which means every string (set) in Σ^+ can be represented as a concatenation of a set in Σ^* and a set in Σ. However, if one forms a set that cannot be defined by a formula but what people would believe as existing, then while the proof itself does not break down, it may be possible that state of matter is inconsistent. This paper explores this possibility.

Additional Key Words and Phrases: Formal language, Zorn's lemma, total order, Kleene star, Kleene plus, computer science, set theory

ACM Reference Format:
M. Kim, 2015. Formal language, intuition, total order, Kleene plus and Zorn's lemma *ACM* V, N, Article A (January YYYY), 2 pages.
DOI: http://dx.doi.org/10.1145/0000000.0000000

1. NON-RIGOROUSLY DEFINED FORMULA, ZORN'S LEMMA AND $\Sigma^+ = \Sigma^*\Sigma$

Let there be a finite character set Σ. Let k be a natural number. Every set(string) in Σ_k is a concatenation of k characters in Σ. Or to formally define inductively,

$$\Sigma_0 = \{\varepsilon\},\ \Sigma_1 = \Sigma$$

where ε represents an empty string,

$$\Sigma_k = \{ab | a \in \Sigma_{k-1},\ b \in \Sigma\}$$

where ab represents concatenation of string a and b. A string is defined rigorously as following: for ith character ($i \in \mathbb{N}^+$) c of string a, it is written as list (c, i). Thus, if $\Sigma = \{0, 1, 2\}$ and string is 011, then it is represented by a set $\{(0, 1), (1, 2), (1, 3)\}$. However, for easy construction of sets, I will forgo the rigorous construction.
$\Sigma^+ = \Sigma^*\Sigma$ is usually proved the following way:

$$\Sigma^+ = \bigcup_{k\geq 1} \Sigma_k = \bigcup_{k\geq 0} \Sigma_{k+1} = \bigcup_{k\geq 0} \Sigma_k\Sigma = \left(\bigcup_{k\geq 0} \Sigma_k\right) \Sigma = \Sigma^*\Sigma$$

Let me deviate, however, and think of building the following set.

At stage 1, start with any arbitrary character $c \in \Sigma$. $c = s_1$.
At stage i, $s_i = s_{i-1}d$ with any arbitrary $d \in \Sigma$.
At stage ω_0, the first infinite ordinal, $s_{\omega_0} = \bigcup_{k\in\mathbb{N}^+} s_k$.
Axiom of union in Zermelo-Fraenkel set theory guarantees that such an arbitrary union must exist.
Now note that s_{ω_0} has been constructed in lexicographical order $\leq$. And $s_i \leq s_{i+1}$, $s_i \leq s_i$.
For any $k \in \mathbb{N}^+$, $s_k \leq s_{\omega_0}$. This does not depend on whether $s_{\omega_0} = s_k$ for some $k \in \mathbb{N}^+$ or not. (However, see next:)
s_{ω_0} cannot be of finite length; otherwise, s_{ω_0} would already have been formed at stage

DOI: http://dx.doi.org/10.1145/0000000.0000000

$k \in \mathbb{N}^+$.
Thus, one can think of extending Σ^* by adding all possible s_{ω_0} from character set Σ. Let us refer to this extended set as Π.

Looking back at how s_{ω_0} is constructed, before reaching stage ω_0, the number of possible s_k at any stage k is finite. Thus, at stage ω_0, the number of all possible s_{ω_0} would be countably infinite (cardinality $\aleph_0$). As the cardinality of Σ^* is countably infinite, Π has cardinality of $\aleph_0$.
But one can map each string s in Π with the power set $P(\mathbb{N})$ of $\mathbb{N}$ by following:

$$s \to \{x \in \mathbb{N} | s_x = a\}$$

where s_x represents xth character of string s. This map is bijective map when $|\Sigma| = 2$, and thus suggests that Π is uncountably infinite.

What if one assumes for whatever reason that s_{ω_0} may not be of infinite length but of finite length? Lexicographical order $\leq$ still follows. If one assumes that each s_{ω_0} is an upper bound in Π for the corresponding totally ordered subset, then by Zorn's lemma, which one can assume if one extends Zermelo-Fraenkel set theory with axiom of choice, states that Π should have at least one maximal element. This result is in contradiction to the generally-held statement that there exists no longest element in Σ^*.
Of course the assumption that s_{ω_0}'s are upper bounds in Π for totally ordered subsets is just an assumption, and may be dropped. However, the matter gets worse if one assumes that $\Pi - \varepsilon = \Sigma^*\Sigma$. This is a plausible assumption, if one assumes that s_ω is not of infinite length, and any s_{ω_0} would likely terminate with some character c. This however necessarily makes true that s_{ω_0} is an upper bound in Π for the corresponding chain, because it now has a terminating character. Thus, Zorn's lemma kicks in, and as $\Sigma^*\Sigma = \Sigma^+$, Σ^+ has at least one maximal element, implying that there exists at least one longest string.

2. RESOLUTIONS

The construction of s_{ω_0} relies on the fact that each character appended is arbitrary. Thus, the set of all possible s_{ω_0}'s is defined not by the formula, but semi-arbitrarily. This allows one to say that Zermelo-Fraenkel set theory would produce what is underlined above. And this indeed is true. However, while mathematics definitely is not an intuitive science, all sets produced as s_{ω_0}'s should exist intuitively or common-sense-wise.
If the first resolution seems not to be the right way to resolve, one may say that constructed Π is indeed $\Pi = \Sigma^*\Sigma$ but $\Sigma^*\Sigma \neq \Sigma^+$. But this contradicts the proof given above.
In a way, the second resolution has intuitive appeals. After all, there is (or we assume that there is) no longest string in Σ^+. That means that the end character is basically unbounded, and one does not have access to it. But $\Sigma^*\Sigma$ allows us access to the end character.
Now as said before, this "problem" or question already has an answer that this is not set-theoretic prediction. However, small intuition speaks that somehow set-theoretical worlds are more unintuitive than often thought. That is at least my conclusion for now.

Received Preprint; revised Preprint; accepted Preprint