1 Introduction

[The ideas behind this parser can be found in the book "A Mathematical Introduction to Logic," by Herbet B. Enderton, and also below. Here, we present the theoretical foundation behind propositional formulas, and conclude with 6 properties of wffs that the parsing algorithm uses, along with the soon-to-be defined idea of construction sequences, to conclude whether user-entered strings are wffs.]

To begin, we define our alphabet A to be the set $\{a - z, A - Z, ', (,), \tilde{}, \&, |, <, =, >\}$. Note that $a-$z means the 26 lowercase letters from $a - z$ inclusive; an analogous comment applies to $A - Z$. Note that our alphabet also contains the apostrophe symbol.

Now we define our universal set U to be the set of all strings of finite, non-zero length formed using the characters in A.

Let V be the set of propositional variables; that is, V is the subset of U such that every element of V begins with a letter followed by zero or more apostrophes.

We now define:

- $f_{\sim} : U \to U$ to be such that $f(x) = \sim x$
- $f_{=>} : (U \times U) - > U$ is such that $f_{=>}(x, y) = (x => y)$
- $f_{\&} : (U \times U) \to U$ is such that $f_{\&}(x, y) = (x\&y)$
- $f_{||} : (U \times U) \to U$ is such that $f_{||}(x, y) = (x||y)$
- $f_{<=>} : (U \times U) \to U$ is such that $f_{<=>}(x, y) = (x <=> y)$

We will call these the connective functions; $f_{\sim}$ is the unary connective function, and the rest are the binary connective functions.

2 An Intuitive Construction of the Set of Wffs

First, we define a *construction sequence* to be a finite sequence of elements of U, say $(a_1, \dots, a_n)$, where every term a_i is either a propositional variable or is the result of applying one of the connective functions to a term or to terms in the sequence with index/indices less than i. The *length* of a construction sequence is the number of terms in the sequence.

If there is a construction sequence whose last term is a string x, then we say that x *has a construction sequence.*

Now we define:

$F_* =$ set consisting of all elements of U that have construction sequences.

This is our set of wffs.

Observe:

Every term in any construction sequence is a wff.

To see this, observe that any construction sequence of length one has, as its only term, a propositional variable. So any term of any construction sequence of length one is a wff.

Now, simply note that for $n > 1$, each of the $n-1$ strict subsequences (a_1), (a_1, a_2),..., $(a_1, \dots, a_{n-1})$ of a construction sequence $(a_1, \dots, a_n)$ are themselves construction sequences, and that, by definition of F_*, the last terms of construction sequences are wffs, and so $a_1, a_2, ..., a_n$ are all wffs.

Observe that here, it is fairly easy to show that something is a wff: one merely exhibits a construction sequence. It is, in general, not so easy to show that something is not a wff using the definition of F_*; to address this, we provide an alternate, equivalent construction of the set of wffs below:

3 An Equivalent Construction

First, a few more definitions:

If S is a subset of U, and if $x, y \in S$ implies that $\sim x, (x => y), (x\&y), (x||y), (x <=> y)$ are all in S, then we say that S is *closed* under the five functions defined above; for brevity, we will simply say S is *closed under* $\{f\}$.

If $V \subset S \subset U$, and if S is closed under $\{f\}$, then we say that S is a *inductive* set.

We now define:

$F^* =$ The intersection of all inductive sets.

Note that F^* itself is inductive.

To see this: V is obviously a subset of F^*, and if $x, y \in F^*$, then x, y are in every inductive set, and so $\sim x, \ldots, (x <=> y)$ are in every inductive set, and so $\sim x, \ldots, (x <=> y)$ are in F^*. That is, $F^* \supset V$, and F^* is closed under $\{f\}$; therefore, F^* is inductive.

We now show that $F_* = F^*$:

First, we'll show $F_* \subset F^*$:

Suppose $x \in F_*$, and so x has a construction sequence $(a_1, \ldots, x)$. a_1 is a propositional variable, and since F^* is inductive, a_1 is in F^*. Now suppose that terms $1, \ldots, n$ $(n \geq 1)$ of the sequence are in F^*. Since term $n+1$ is either a propositional variable or something formed via application of one of the connective functions to some of the preceding term(s), and since F^* is closed under $\{f\}$, term $n+1$ must be in F^*.

So by induction, $x \in F^*$.

Now we'll show $F^* \subset F_*$.

First, observe that F_* is inductive: $F_* \supset V$, and if $x, y \in F_*$, then they have construction sequences $(a_1, ..., x)$ and $(b_1, ..., y)$, and so $(a_1, \ldots, x, \sim x)$ and $(a_1, \ldots, x, b_1, \ldots, y, g(x, y))$ are themselves construction sequences, where g is any binary connective function. So F_* is closed under $\{f\}$.

So if $x \in F^*$, then x is in every inductive set, and so $x \in F_*$.

From now on, we will refer to the set of wffs using F, with the understanding that

$$F = F_* = F^*$$

4 The Induction Principle

As promised, we will now address the issue of showing that something is not wff:

To do so, we first observe that F "absorbs" any inductive set that is a subset of F:

That is, if S is inductive, and S is a subset of F, then $S = F$.

To see this, we repeat the argument above: since $F =$ the intersection of all inductive sets, then F is a subset of every inductive set, including S, and so $S = F$.

This gives us something very useful: we can specify a property P, form a subset S of F that consists of all the elements of F that have property P, and if we can show that S is a inductive set, then we can conclude that all members of F have property P.

This is very similar to mathematical induction, and so we call this the induction principle.

The induction principle allows us to assert that wffs have the 6 properties below, and that any string that lacks any one of these properties is not a wff. We will give proofs for our claims that wffs have property P_5 and P_6; the other properties can be easily verified.

5 Properties of Wffs

1. Any wff x has property P_1: its number of left parentheses is equal to its number of right parentheses (we will say that x is "balanced").

2. Any wff x has property P_2: it begins with either a letter, the $\sim$ character, or a (.

3. Any wff x has property P_3: If x begins with a letter, then x is a letter followed by zero or more apostrophes.

4. Any wff x has property P_4: If x begins with (, then it ends with).

5. Any wff x has property P_5: If x begins with (, then x has at least one proper initial segment, and any proper initial segment of x is such that its number of left parentheses is greater than its number of right parentheses (we say that the segments are left-heavy).

   Proof:

   Suppose S is the set of all elements of F with property P_5. Then all propositional variables are vacuously in S, and so $S \supset V$. If $x, y \in S$, then $\sim x$ is vacuously in S. Consider $g_C(x, y) = (xCy)$, where C is any binary connective. By inspection, (xCy) must have a proper initial segment. By the fact that x, y are also in F, and by property P_1, x and y are both balanced, and so any proper initial segment is necessariy left-heavy, by virtue of the leftmost left bracket. So $(xCy) \in S$.

   Hence, S is closed under $\{f\}$, and by the induction principle, $S = F$.

6. Any wff x has property P_6: If x begins with $\sim$, then $x = \sim \cdots \sim p' \ldots '$ or $x = \sim \cdots \sim (\ldots)$. Here, p stands for any letter, and $\sim \cdots \sim$ means one or more $\sim$'s and $' \ldots '$ means zero or more apostrophes. (...) means something beginning with '(', ending with), and also something that is balanced and having at least one proper initial segement, and whose proper initial segments are left-heavy.

   Proof:

   Suppose S is the set of all elements of F with property P_6. Then all propositional variables are vacuously in S, and so $S \supset V$. If $x \in S$, then x either begins with $\sim$ or doesn't. If it does, then $x = \sim \cdots \sim p' \ldots '$ or $x = \sim \cdots \sim (\ldots)$, and so $\sim x \in S$. If x doesn't begin with $\sim$, then, by the fact that $x \in F$ as well, and by property P_1 above, x begins with either (or a letter.

   If x begins with a letter p, then, by property P_3, $x = p' \ldots '$. So $\sim x$ is in S. If x begins with (, then by property P_1, P_4, and P_5, $x = (\ldots)$. So $\sim x \in S$.

   For any $x, y \in S$, and any binary connective function g, $g(x, y)$ is vacuously in S.

   So S is closed under $\{f\}$ and is thus inductive. Therefore, $S = F$, and every element of F has property P_6