Integrating pattern data types with Unicon string scanning

John H. Goettsche
Dept. of Computer Science
University of Idaho

July 10, 2014

Abstract

1 Introduction

Editing, manipulating and analyzing text are common tasks performed by programming languages. How a programming language searches for patterns within a body of text is an important area of computer science affecting the effectiveness and efficiency in scanning text.

The first section of this paper will review history and development of the scanning frameworks that originated from the SNOBOL4-Icon programming languages.

2 Background

2.1 SNOBOL4

SNOBOL4 was developed by Bell Telephone Laboratories in 1962. Searching for a desired pattern within a string of characters is one of its basic operations. This pattern matching could be as simple as a single character or a set of characters in a particular order, or it can be a complex arrangement with alternative character sets. A pattern data type was used enabling the user to define and store patterns as variables. [1]

The pattern matching statement in a SNOBOL4 program is in the form of a subject followed by at least one space then the pattern.

```
SUBJECT  PATTERN
```

In the above line, SUBJECT would be scanned to see if it contains the contents of PATTERN, if it succeeds, then a substring of SUBJECT that fits PATTERN

would be returned. In the anchored mode the pattern would have to begin its match at the first character of the subject string; in the non-anchored mode the pattern could start at any character in the string.

2.2 Icon-Unicon

Icon's string scanning framework used an ...

2.3 Sudarshan Gaikaiwari's Implementation of SNOBOL4 Patterns

3 Design Considerations

4 Implementation

References

[1] Ralph E Griswold, Ivan Paul Polonsky, and JF Poage. The snobol4 programming language. 1971.