

PA193 project part 2 – Ron Rivest's S-Expressions

- Specification at <http://people.csail.mit.edu/rivest/Sexp.txt>
- Representation of nested structures of octet strings (byte arrays).
- S-Expression is either an octet string or a finite list of simpler S-Expressions.
- Octet strings have a number of possible representations, including: symbol (text identifier containing a specific subset of characters), quoted text, base16 encoding, base64 encoding, length followed by verbatim byte string.
- Except for symbol, every octet-string format can be prefixed by length of the string in decimal notation.
- Looks like a text format, but is not – text-binary hybrid.
- Does not have regular lexical tokens.
- Originally proposed as a format for cryptographic data. Safety of the implementation would have been critical for applications, if the format ever attained widespread use.

Parser

- Similar to a recursive descent parser.
- Uses one-byte look-ahead to determine what is next:
 - Start of list.
 - Length of an octet string, followed by octet string.
 - Specific octet string format.
- Parses a corresponding S-Expression.
- Skips whitespace whenever allowed.
- List parsing recursively parses S-Expression until end-of-list is found.
- Approximately 600 LOC total.
- For convenience, parser accepts more generic form of input than specified, in some cases.