Regular expressions can be used to

- Specify a set of strings
- Search a text for patterns
- Generate responses for a simple discourse agent
- Specify an entire human language (like French)
- Write recursive algorithms
- An **alphabet** is a set of symbols, $\Sigma$.
- A **string** over an alphabet is a sequence of symbols from that alphabet. $\Sigma^*$ is the set of all strings over alphabet $\Sigma$.
- A **language** over an alphabet is a set of strings, that is, a subset of $\Sigma^*$.

- **Regular expressions** constitute a system for specifying languages. (J&M, “a language for specifying text search strings”, pg 3.). An individual regular expression denotes a language, that is, a set of strings.
\[ \emptyset \quad \text{the empty set of strings} \\
\varepsilon \quad \text{the set containing the empty string, } \{""\} \\
a \quad \text{the set containing only the string with only } a, \\
\text{for some } a \in \Sigma, \{"a"\} \\
rs \quad \text{the set of strings made from concatenating strings from } r \text{ and } s, \\
\{x + y \mid x \in r \land y \in s\}, \text{ for some regular expressions } r \text{ and } s \\
r|s \quad \text{the set of strings from } r \text{ or } s, r \cup s \\
\text{for some regular expressions } r \text{ and } s \\
r^* \quad \text{the set of strings made from concatenating 0 or more strings from } r \\
\text{for some regular expression } r \]
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Meaning</th>
<th>Equivalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>[abc]</td>
<td>One occurrence of any of these symbols</td>
<td>(a</td>
</tr>
<tr>
<td>[a–c]</td>
<td>One occurrence of any symbol in this range</td>
<td>(a</td>
</tr>
<tr>
<td>r?</td>
<td>Optionally an occurrence of a string defined by ( r )</td>
<td>( (r</td>
</tr>
<tr>
<td>( r^5 )</td>
<td>5 occurrences of a string defined by ( r )</td>
<td>( rrrrr )</td>
</tr>
<tr>
<td>( r^{3,5} )</td>
<td>Between 3 and 5 occurrences of a string defined by ( r )</td>
<td>( rrr</td>
</tr>
<tr>
<td>( r^+ )</td>
<td>One or more occurrences of a string defined by ( r )</td>
<td>( rr^* )</td>
</tr>
</tbody>
</table>
DNA sequences: (A|C|G|T)*.

Identifiers: ( (′ | ε) [A-Za-z] [A-Za-z0-9_] ) | _.


US Postal Addresses: [0-9]+ [NSEW]⁰⁻² [A-Z] [a-z]* (St|Ave|Rd|Ln|Dr|Blvd), ([A-Z][a-z]*)*, [A-Z]²[0-9]⁵.
Lord, you have been our dwelling place in all generations.