

# Newbie Shell Design Document

## Executive Summary

Newbie is a modern, user-friendly Linux shell designed to replace the cryptic syntax and hostile user experience of traditional shells with natural language commands and predictable behavior. Built in Rust with a threaded architecture, newbie leverages existing GNU tools while providing an intuitive interface accessible to beginners and productive for experienced users.

## Core Design Philosophy

### Natural Language Over Cryptic Abbreviations

- Commands use readable English words: find, show, copy, remove
- Syntax follows natural language patterns: find error in logs.txt
- No arbitrary abbreviations requiring memorization

### Predictable Behavior

- Commands do exactly what they say, nothing more or less
- No "smart" behavior that changes based on context
- Consistent output formatting across all commands
- Same input always produces same output
- **Indentation-based structure:** Uses Python-style whitespace indentation instead of brackets, semicolons, or other punctuation clutter

### Minimal Escaping and No String Delimiters

- **No string delimiters:** Eliminates the primary source of shell scripting pain - no quotes, no escaping hell
- **Universal `&` prefix rule:** Any keyword that affects command parsing gets `&` prefix; escape literal versions with `\&`
- **Context-aware parsing:** Different command contexts (if, find, copy) have their own parsing rules while maintaining universal escaping
- **Four foundational patterns:** `&` prefix system, `&start...&end` patterns, `show` with modifiers, and OOP dot notation cover all general cases

## Explicit Output Control

- All display output requires explicit show command unless followed by &raw
- Silent operation by default for scriptability
- Human-readable formatting applied automatically with show unless followed by &raw

## Pattern Language: Readable Alternative to Regular Expressions

### The Regular Expression Problem

Regular expressions are ubiquitous in Unix shells but suffer from fundamental usability issues:

- **Cryptic, unreadable syntax:** `^[A-Z]+[0-9]+\.txt$` is meaningless to most users
- **Poor debugging capabilities:** When regex fails, error messages are unhelpful
- **Single-threaded performance limitations:** Traditional regex engines don't utilize modern multi-core systems
- **Backtracking complexity:** Regex engines often need to look ahead or backtrack, preventing efficient streaming
- **Excessive escaping requirements:** Special characters must be escaped differently in various contexts

### Solution: Left-to-Right Streaming Pattern Language

Newbie implements a pattern language designed for efficient left-to-right processing:

#### Streaming Architecture Advantage

The `&start...&end` format enables **incremental matching** as data streams in:

- **No backtracking required:** Patterns are evaluated left-to-right as characters arrive
- **Real-time processing:** Large files can be processed without buffering entire contents
- **Multi-threaded performance:** Reader/worker/writer threads process data continuously
- **Memory efficient:** Only small working buffers needed, not entire file contents

Example streaming pattern: `find &start error &numbers &end in huge_logfile.txt`

- Processes character by character as file is read
- Immediately identifies matches without storing entire file in memory
- Scales to arbitrarily large files

### Natural Language Pattern Syntax

Newbie implements readable patterns using natural English tokens prefixed with `&`:

## Basic Elements

`&start` and `&end` - Beginning and end anchors (equivalent to regex `^` and `$`)

`&text` - Any characters (equivalent to regex `.`)

`&letters` - Any letters, case-insensitive (equivalent to `[A-Za-z]*`)

`&upperletters` - Uppercase letters only (equivalent to `[A-Z]*`)

`&lowerletters` - Lowercase letters only (equivalent to `[a-z]*`)

`&numbers` - Numeric digits (equivalent to `[0-9]*`)

## Quantified Matching

All character classes support optional numeric quantifiers for exact matching:

`&text5` - Exactly 5 text characters

`&letters3` - Exactly 3 letters

`&upperletters2` - Exactly 2 uppercase letters

`&numbers4` - Exactly 4 numbers

## Real-World Examples

**Traditional regex** (unreadable): `^[A-Z]+[0-9]+\..txt$`

**Newbie pattern** (readable): `&start &upperletters &numbers .txt &end`

**Traditional regex:** `error.*[0-9]+`

**Newbie pattern:** `&start error &text &numbers &end`

**Complex literal text** (impossible to escape cleanly in regex):

`find &start he said, "copy the file to C:\hosts". I don't know why &end in file.txt`

## Wildcards

`&*` - Multiple character wildcard

`&?` - Single character wildcard

## Anchoring

&start - Beginning of text

&end - End of text

&start = foo - Text must start with "foo" (equivalent to regex ^foo)

&end = bar - Text must end with "bar" (equivalent to regex bar\$)

&start = [ERROR] &end = process complete - Text must start with "[ERROR]" and end with "process complete"

&start = HTTP/1.1 &numbers3 &end = Connection: close - HTTP response pattern with specific start/end boundaries

## Logic Operators

&or - Alternation (A or B)

&not - Negation

&maybe - Optional element

## Proximity Matching

&within 5 words of

&within 3 characters of

## Mode Detection and Integration

The parser automatically detects search mode based on content:

- **Literal mode:** No ampersands present - simple substring search
- **Pattern mode:** Ampersands present - full pattern language active

### Examples:

find error in logs.txt

- Literal substring search for "error"

find &start error &numbers &end in logs.txt

- Pattern search for "error" followed by numbers at end of line

find &start he said, "copy the file to C:\hosts". I don't know why &end in file.txt

- Complex literal text with quotes, backslashes, and punctuation - no escaping needed

## Variable embedding in patterns:

```
&v.error_code = 404  
find &start error &v.error_code &end &in access.log
```

- Dynamic pattern construction

## Command Architecture

### Unified Commands

Traditional shells scatter related functionality across multiple tools with different syntaxes. Newbie unifies related operations under intuitive command names.

### Find Command

Replaces ls, grep, and find with context-aware behavior. **Directory vs File distinction:** Uses trailing slash to clearly identify directories - config/ means search within the directory, while config.txt means search within the specific file.

### File listing

```
find *.txt  
find *.log &in /var/log/
```

### Content search (literal mode) - no delimiters needed for complex strings

```
find error &in logs.txt  
find configuration issue &in config/  
find The user said "hello world" and received 'no response' from server &in application_logs/
```

### Content search (pattern mode - requires &start and &end)

```
find &start error &numbers &end &in logs.txt  
find &start &upperletters3 &numbers4 &end &in part_codes.txt
```

### Show Command

Universal display command with automatic human-readable formatting and composable modifiers:

```
show file.txt          # Paged display (less equivalent)
show file.txt &raw      # Raw output (cat equivalent)
show file.txt &formatted # With syntax highlighting
show file.txt &numbered # With line numbers
show file.txt &first 20 &lines # First 20 lines (head equivalent)
show file.txt &last 20 &lines # Last 20 lines (tail equivalent)
show file.txt &first 1000 &chars # First 1000 characters
show file.txt &last 1000 &chars # Last 1000 characters
```

Modifiers can be combined for powerful display options:

```
show file.txt &raw &first 100 &lines      # Raw first 100 lines for piping
show file.txt &formatted &numbered &last 50 &lines # Formatted last 50 lines with numbers
show &system.memory                      # System memory information
show &process.firefox                    # Process information
```

## File Operations

Intuitive syntax matching natural language with trailing slash convention for directories:

```
copy file.txt &to backup/
move oldname &to newname
remove file.txt
remove directory/
```

**Directory vs File distinction:** Trailing slash clearly identifies directories:

- `if /etc/bashrc then` - file exists
- `if ~/.bashrc.d/ then` - directory exists
- `if ../config/ then` - parent directory exists
- `if / then` - root directory exists

## Text Processing

Readable replacement syntax:

```
replace foo &with bar &in file.txt
replace old text &with new text &in *.txt
replace &start error &numbers &end &with FIXED &in logs.txt
```

## Output Operations

Natural language output redirection:

```
find results &to filename.txt    # write output to file (overwrite)
find results append &to filename.txt  # add output to end of file
```

## System Information

Simple commands for common system queries:

```
space    # Filesystem usage (defaults to home directory)
space /etc # Specific path
memory   # Memory information
```

## Administrative Operations

Clear privilege escalation:

```
admin space /etc
admin remove /system/file
```

## Arithmetic and String Operations

### Arithmetic Operators

All arithmetic operators use `&` prefix to avoid conflicts with natural data:

- `&+` for addition/concatenation
- `&-` for subtraction
- `&*` for multiplication
- `&/` for division

This approach recognizes that arithmetic symbols appear frequently in real data (file paths, URLs, mathematical expressions, log files) but `&+` sequences are extremely rare.

Examples:

```
total = price &* quantity
filename = base &+ extension
result = memory &- used_memory
```

## String Concatenation

Simple concatenation using `&+` operator:

```
new_name = processed_ &+ file
full_path = directory &+ / &+ filename
```

## Wildcard System

### Problem

Traditional wildcards (\* and ?) conflict with the goal of making these characters searchable without escaping.

### Solution

Use unlikely two-character combinations:

- `&*` for multi-character wildcards
- `&?` for single-character wildcards

This allows literal \* and ? in search terms while preserving wildcard functionality:

```
find &*.txt # Wildcard matching
find *.txt  # Literal asterisk search
```

Escaping only needed for the rare cases of literal `&*` or `&?` in content: `\&*`

## Escaping Rules

### Universal Escaping Principle

**Single rule for all contexts:** Any keyword that affects command parsing gets an `&` prefix. To search for the literal version of a parsing keyword, escape it with a backslash.

Examples across different commands:



```
find text with \&in literal &in file.txt  
copy file\&to.txt &to destination/  
replace old text \&with literal &with new text &in file.txt
```

## Context-Aware Parsing

Different command contexts have their own parsing rules while maintaining the universal escaping principle:

- **If context:** Tests existence or conditions
- **Find context:** Searches files or content
- **Copy context:** Handles source and destination
- **For context:** Iterates over collections

This eliminates ambiguity - a file named "then" doesn't conflict with the `then` keyword because context determines meaning.

## Programming Language Features

### Variables and Arithmetic

Clean syntax without shell quoting complexities:

```
&v.x = 2 &+ 2  
&v.result = memory  
&v.total = &v.price &* (1 &+ tax_rate)  
show &v.total
```

### Variable Scoping and Syntax

All variables and properties use `&namespace.` prefix for unambiguous identification in delimiter-free parsing:

- **Local variables:** `&v.` prefix for user-defined variables
  - `&v.file = data.txt`
  - `&v.count = 0`
  - `&v.result = calculation`
- **System environment variables:** `&system.` prefix
  - `&system.path = /usr/bin:/bin`

- `&system.home = /home/username`
- `&system.user = username`
- `&system.shell = /usr/bin/newbie`
- **Global variables:** `&global.` prefix for cross-session persistence
  - `&global.config_file = ~/.newbie/config`
  - `&global.default_editor = nano`
- **System properties:** `&system.` namespace extends to structured data
  - `&system.memory.free`
  - `&system.cpu.load`
  - `&system.disk.root.free`
- **Process information:** `&process.` namespace
  - `&process.firefox.status`
  - `&process.123.memory`
- **Configuration data:** `&config.` namespace
  - `&config.database.host`
  - `&config.database.port`
- **Network information:** `&network.` namespace
  - `&network.interface.eth0.ip`
  - `&network.connection.ssh.active`

The consistent `&namespace.property` pattern enables embedding variables in any context without delimiters:

```
&v.error_code = 404
find &start error &v.error_code &end &in access.log
show &system.memory.free
if &process.nginx.status equals running then
```

This approach creates a universal, discoverable interface to all system data while maintaining the delimiter-free design philosophy.

## Control Flow

**Indentation-based structure** (like Python): Uses whitespace indentation to define code blocks, eliminating brackets, semicolons, and other punctuation clutter.

## Conditional Statements

**Filesystem path conventions:** Trailing slash distinguishes directories from files, supporting all standard path patterns:

```
if /etc/bashrc then
  load /etc/bashrc
end

if ~/.bashrc.d/ then
  for file in ~/.bashrc.d/&*
    if file then
      load file
    end
  end
end

if / then          # root directory
if ../config/ then  # parent directory
if ../../shared/ then  # multiple parent levels
if ./local/ then    # current directory (explicit)
if subdirectory/ then  # current directory (implicit)
```

## Pattern Matching in Conditionals

```
for file in &*.txt
  if file matches &start &upperletters &numbers .txt &end then
    move file &to processed_ &+ file
  end
end
```

## Loops

```

for file in &*.txt
  show Processing: &+ file
  size = get_file_size file
  if size greater than 1MB then
    compress file
  end
end

do while tasks_remaining greater than 0
  process_next_task
  tasks_remaining = tasks_remaining &- 1
end

```

## Comparison Operators

Natural language and symbolic operators with **(&)** prefix:

- *less than* / **&<**, *greater than* / **&>**, *equals* / **&=**
- *greater than or equal* / **&>=**, *less than or equal* / **&<=**
- *contains*, *starts with*, *ends with*
- *matches* for pattern matching

## File and Directory Properties

Object-oriented property access for file system information:

- *filename.size* - file size in bytes
- *filename.lines* - line count
- *filename.chars* - character count
- *filename.modified* - last modified time
- *filename.permissions* - file permissions
- *directory/.count* - number of items in directory

Context disambiguates between filenames and properties:

```

if /boot/uboot/uboot.env.count &> 1 then
if logfile.lines &= 0 then
if directory/.count &< 5 then

```

## Debugging Support

`show` statements positioned at leftmost column within current indentation level for easy visual scanning:

```
&v.x = get_input
show Input received: &+ &v.x
if &v.x greater than 100 then
show Large value processing
  &v.result = complex_calculation &v.x
show Calculation result: &+ &v.result
end
```

## Implementation Architecture

### Technology Stack

- **Language:** Rust for memory safety and performance
- **Threading:** Reader/worker/writer pattern across all components
- **Tool Integration:** FFI bindings to GNU utilities via existing C libraries where appropriate; many core utilities like `xargs`, `sync`, `sort`, `uniq` work well as-is
- **Parsing:** Adventure-game-style pattern matching for natural language with context-aware grammar

### Performance Strategy

#### Two-phase execution model:

1. **Parse phase:** Natural language commands parsed once into efficient Rust operations
2. **Runtime phase:** Compiled operations execute at native speed

**Left-to-right streaming architecture:** The natural language format inherently supports streaming processing. Patterns like `find &start error &numbers &end &in logs.txt` can be matched incrementally as data arrives, without requiring backtracking or look-ahead. This enables real-time processing of large files without buffering entire contents.

Multi-threaded architecture with three-thread pattern:

- **Reader thread:** Stream input data character by character
- **Worker thread:** Process operations incrementally using left-to-right matching
- **Writer thread:** Format and output results as they become available

This streaming approach maximizes throughput on modern multi-core systems while maintaining low memory usage, and provides compatibility with existing GNU tool reliability.

## Core Utility Philosophy

Newbie enhances rather than replaces:

- **Transform the painful parts:** Complex find operations, control flow, variable handling, pattern matching
- **Preserve what works:** Tools like `sort`, `uniq`, `xargs`, `sync` that are already clear and efficient
- **Unify scattered functionality:** `show` replaces `cat`, `less`, `head`, `tail` with natural language modifiers
- **Enable integration:** Clean piping between newbie natural language and traditional utilities

## GNU Tool Integration

Rather than reimplementing functionality, newbie provides natural language interfaces to battle-tested GNU utilities:

- Leverage existing reliability and feature completeness
- Focus innovation on user experience rather than core functionality
- Use FFI and existing Rust crates (readline, etc.) for integration

## Pipeline Operations

Support both traditional and natural syntax:

- Traditional: `command1 | command2`
- Natural: `command1 into command2`

Example:

```
find error &in logs.txt into show
find &upperletters3&numbers4 &in data.txt into count into show
```

## Configuration Philosophy

Replace bash's cryptic configuration syntax with human-readable alternatives:

### Traditional Bash Configuration

```
bash
```

```

PS1="\[\033[01;32m\]\u@\h\[\033[00m\]:\[\033[01;34m\]\w\[\033[00m\]\$ '
alias ll='ls -aF'
alias la='ls -A'
alias l='ls -CF'
export PATH="$HOME/bin:$PATH"

if ! [[ "$PATH" =~ "$HOME/.local/bin:$HOME/bin:" ]]; then
    PATH="$HOME/.local/bin:$HOME/bin:$PATH"
fi
export PATH

# User specific aliases and functions
if [ -d ~/.bashrc.d ]; then
    for rc in ~/.bashrc.d/*; do
        if [ -f "$rc" ]; then
            . "$rc"
        fi
    done
fi
unset rc

# >>> conda initialize >>>
# !! Contents within this block are managed by 'conda init' !!
__conda_setup="$('/usr/bin/conda' 'shell.bash' 'hook' 2> /dev/null)"
if [ $? -eq 0 ]; then
    eval "$__conda_setup"
else
    if [ -f "/usr/etc/profile.d/conda.sh" ]; then
        . "/usr/etc/profile.d/conda.sh"
    else
        export PATH="/usr/bin:$PATH"
    fi
fi
unset __conda_setup
# <<< conda initialize <<<
. "$HOME/.cargo/env"

```

## Newbie Configuration

prompt:

user: green bold

host: green bold

path: blue bold

symbol: default

shortcuts:

ll = find all &with details &with types

la = find all hidden

l = find all &with columns

```
if &system.path not contains ~/.local/bin: &+ ~/.bin: then
  &system.path = ~/.local/bin: &+ ~/.bin: &+ &system.path
end
```

# User specific aliases and functions

if ~/.bashrc.d/ then

for &v.file in ~/.bashrc.d/&\*

if &v.file then

load &v.file

end

end

end

# >>> conda initialize >>>

&v.conda\_setup = run /usr/bin/conda shell.newbie hook

if &v.conda\_setup succeeded then

eval &v.conda\_setup

else

if /usr/etc/profile.d/conda.sh then

load /usr/etc/profile.d/conda.sh

else

&system.path = /usr/bin: &+ &system.path

end

end

# <<< conda initialize <<<

load ~/.cargo/env

## Migration Strategy

### Bash-to-Newbie Translator

Develop translation tool to convert existing bash scripts to readable newbie equivalents:

- Focus on functional intent rather than syntax replication



- Handle common patterns and idioms
- Identify newbie feature gaps requiring development
- Generate readable code that accomplishes same goals

## Compatibility

- Support traditional operators where unambiguous (`|`, `~`)
- Provide natural language alternatives as primary interface
- Allow gradual migration from existing workflows

## Error Handling

Leverage Rust's Result types for better error messages:

- Context-aware error descriptions
- Specific suggestions for common mistakes
- Clear indication of where parsing failed
- Helpful guidance for escaping conflicts

Examples:

Error: Found '&with' in search text - did you mean to search for literal '&with'?

Try: replace old text \&with literal &with new text &in file.txt

Error: Found '&to' in filename - did you mean literal '&to'?

Try: copy file\&to.txt &to destination/

## Design Rationale: Why No String Delimiters?

### The Core Problem

String delimiters are the primary source of shell scripting pain, wasting countless hours on:

- Quote escaping nightmares: `ssh host "grep 'pattern with \"quotes\" file"`
- Variable quoting bugs: `rm $filename` fails when filename contains spaces
- Nested delimiter hell in JSON/XML processing
- Database query building with multiple escaping layers: SQL record `"The Matrix"` becomes a syntax nightmare
- The eternal "single vs double quotes" confusion

## The Solution: Structural Parsing

Newbie eliminates delimiters entirely by using structural parsing:

- **No quotes needed:** `find error message &in logs.txt`
- **No escaping common characters:** Arithmetic symbols, spaces, punctuation are literal
- **Context-aware grammar:** Command structure determines parsing, not delimiters
- **Rare conflicts handled simply:** `\&in` for literal `&in` when it conflicts with parsing keywords

## Real-World Impact

This design emerged from practical experience with SQL injection and escaping problems. The approach means:

- Common data (with spaces, quotes, symbols) needs no escaping
- Only truly rare conflicts (like literal `&in` text) require simple `\&` escaping
- Mental energy focuses on logic, not syntax mechanics
- The `&` key gets more use, but eliminates far more complex escaping scenarios

## Success Metrics

### User Experience

- Reduced learning curve for shell newcomers
- Faster task completion for common operations
- Fewer syntax errors and debugging sessions
- Improved script maintainability

### Performance

- Faster text processing through multi-threading
- Competitive performance with traditional tools through parse-once, execute-fast model
- Efficient resource usage

### Adoption

- Standalone pattern matching tool adoption
- Integration into Linux distributions
- Community contribution and extension

## Future Considerations

### Pattern Language Extensions

- Additional character classes as needed
- More sophisticated proximity matching
- Performance optimizations for complex patterns

### Command Set Expansion

- Network operations with natural syntax
- Archive manipulation commands
- Process management improvements
- **Display command unification:** `show` replaces `cat`, `less`, `head`, `tail` with composable natural language modifiers
- **Core utility integration:** Leverage existing tools like `xargs`, `sync`, `sort`, `uniq` that already work well

### Integration Opportunities

- IDE and editor integration for syntax highlighting
- Shell completion systems
- Documentation and tutorial generation

## Conclusion

Newbie represents a fundamental rethinking of shell design, prioritizing human comprehension and modern computing capabilities over historical constraints. By combining natural language interfaces with threaded performance and reliable GNU tool integration, newbie can make shell computing accessible to broader audiences while maintaining the power needed for advanced use cases.

The design eliminates major pain points of traditional shells - cryptic syntax, hostile error messages, poor performance, and maintenance difficulties - while preserving the essential functionality that makes shells powerful tools for system administration and automation.

The elimination of string delimiters alone represents a paradigm shift that could save developers countless hours of syntax wrestling, allowing them to focus on solving actual problems rather than battling quote escaping mechanics.

