Gemini Context Extension

This extension provides two essential tools for managing your Gemini CLI usage:

Available Tools

1. track_context_usage

Analyzes your context window usage and shows:
- Total tokens used vs available
- Breakdown by component (system, tools, MCP servers, extensions, context files)
- Optimization recommendations

Usage modes:
- `compact`: Quick overview
- `standard`: Detailed breakdown (default)
- `detailed`: Full analysis with recommendations

2. estimate_api_cost

Estimates API costs based on your current context and usage:
- Per-request cost calculation
- Cost comparison across different Gemini models
- Budget planning for multiple requests

Parameters:
- `model`: Model name (default: gemini-2.0-flash-exp)
- `requestCount`: Number of requests to estimate (default: 1)

Best Practices

When working with large contexts:
1. Use `track_context_usage` regularly to monitor token usage
2. Review recommendations in detailed mode
3. Compare costs across models using `estimate_api_cost`
4. Consider disabling unused MCP servers to reduce context

Tips

- Context files (GEMINI.md) are discovered hierarchically from your working directory up to project root
- MCP servers contribute significantly to context size (~5k tokens per server)
- Extensions with context files add to your baseline token usage