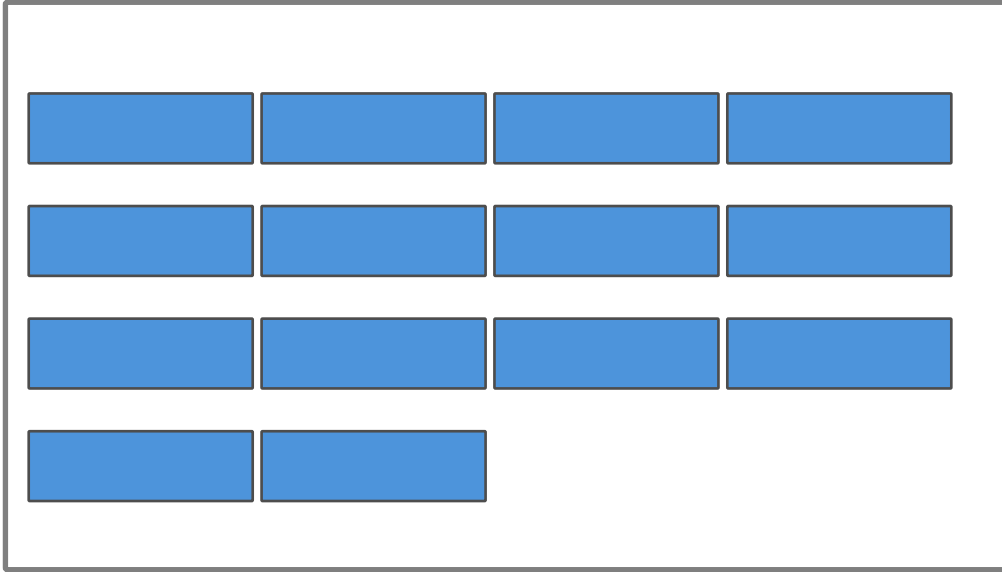
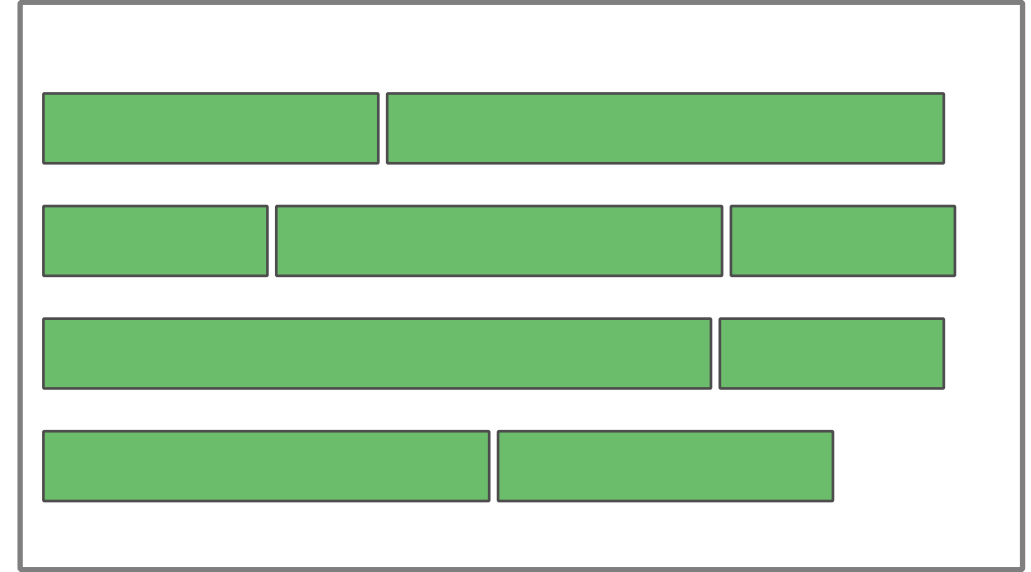


Fixed-Size Chunking



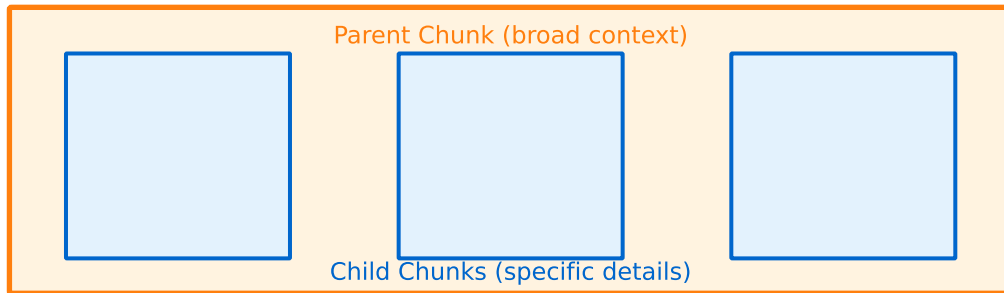
Simple: Split every N tokens (e.g., 512)

Semantic Chunking



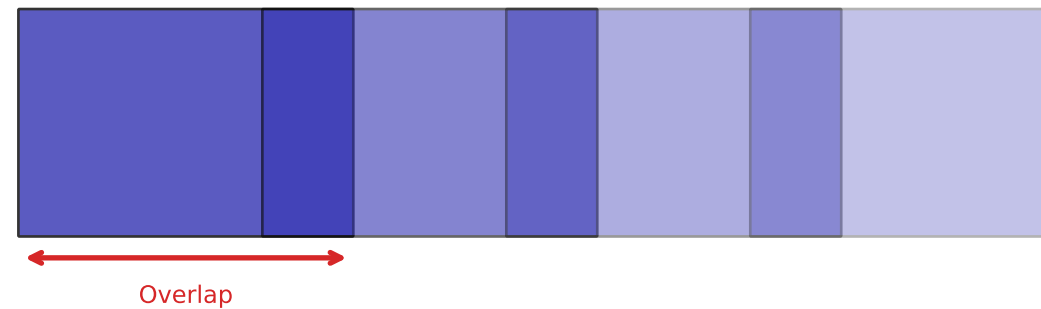
Split at paragraph/section boundaries

Hierarchical Chunking



Multi-level: Query routes to appropriate granularity

Sliding Window



Overlapping windows: No info lost at boundaries