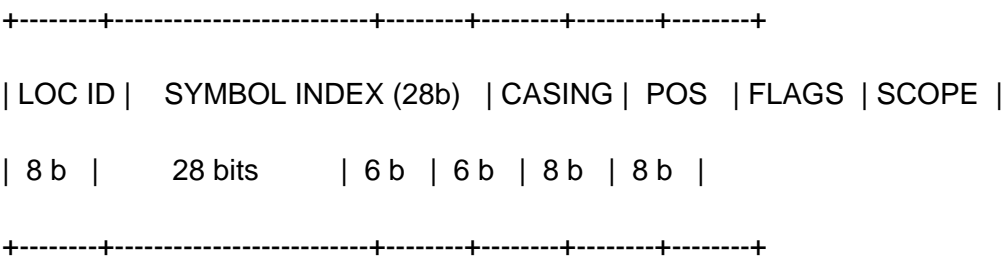


ConC.GPT 64-bit Symbol Format

Overview

The ConC.GPT 64-bit symbol format compresses a word or phrase, along with metadata, into a single 8-byte (64-bit) unit. It is designed for GPT-based workflows, offering language localization, fast decoding, and rich semantic tagging.

Bitfield Layout



Field Descriptions

- LOC ID (8 bits): Language/region index (supports up to 256 dictionaries).
- SYMBOL INDEX (28 bits): Main dictionary lookup (supports ~268M entries per language).
- CASING (6 bits): Encodes case format (e.g. lowercase, UPPER, Title Case).
- POS (6 bits): Part of Speech or token type (noun, verb, emoji, etc.).
- FLAGS (8 bits): Customizable flags (sarcasm, emotion, emphasis, etc.).
- SCOPE (8 bits): Token grouping or contextual association (e.g., sentence/phrase ID).

Key Benefits

- Fixed-length (64-bit) format is fast and hardware-aligned.

ConC.GPT 64-bit Symbol Format

- Supports multilingual and domain-specific dictionaries.
- Allows per-token metadata for LLM interfacing.
- Efficiently compresses large texts and codebases.