**RAG知识库构建细则**

**1. 中文知识库整理规范**

**1.1 文本长度要求**

* 每个三号字体标题下的内容不应超过3500个字符（约1750个中文词）
* 标题本身不计入字符限制

**1.2 结构要求**

* 使用三号字体作为主要分割标题
* 每个三号字体标题下可以包含一个相关图片

**1.3 语义要求**

* 确保每个段落都围绕一个中心主题
* 使用简洁明了的语言，避免冗余
* 定义专业术语，确保内容对非专业人士也易于理解
* 使用标准普通话，避免方言或俚语

**1.4 图片要求**

* 每个图片应与其所属标题直接相关
* 图片分辨率建议不低于300dpi
* 图片格式推荐使用PNG或JPG
* 为每个图片提供简短的描述性说明

**2. 日本語ナレッジベース整理規範**

**2.1 テキスト長さの要件**

* 3号フォントの見出し以下のコンテンツは7000文字（約2500語）を超えないこと
* 見出し自体は文字制限にカウントしない

**2.2 構造要件**

* 3号フォントを主な区切り見出しとして使用
* 各3号フォントの見出しの下に関連画像を1つ含めることができる

**2.3 意味的要件**

* 各段落が一つの中心的なトピックに焦点を当てていることを確認
* 簡潔明瞭な言語を使用し、冗長さを避ける
* 専門用語を定義し、非専門家にも理解しやすい内容にする
* 標準的な日本語を使用し、方言や俗語を避ける

**2.4 画像要件**

* 各画像はそれが属する見出しと直接関連していること
* 画像解像度は300dpi以上を推奨
* 画像フォーマットはPNGまたはJPGを推奨
* 各画像に簡潔な説明文を付ける

**3. English Knowledge Base Organization Guidelines**

**3.1 Text Length Requirements**

* Content under each size 3 font heading should not exceed 5000 words (approximately 25000 characters)
* The headings themselves do not count towards the character limit

**3.2 Structure Requirements**

* Use size 3 font as the main dividing headers
* Each size 3 font header may include one relevant image

**3.3 Semantic Requirements**

* Ensure each paragraph focuses on a central topic
* Use clear and concise language, avoiding redundancy
* Define technical terms to ensure content is accessible to non-experts
* Use standard English, avoiding colloquialisms or regional expressions

**3.4 Image Requirements**

* Each image should be directly relevant to its associated header
* Recommended image resolution of at least 300dpi
* Preferred image formats are PNG or JPG
* Provide a brief descriptive caption for each image