

ニコニコ大百科データセット 近日中に**10年ぶりに最新化**予定

情報学研究データリポジトリで公開中のニコニコデータセット内の「ニコニコ大百科データ」を10年ぶりに更新します。

約51万ページの記事データがあり、LLMの学習等にも利用可能です。特にサブカルチャー文脈の用語に強い日本語のコーパスとして利用できます。



ニコニコ大百科とは

ニコニコ大百科は、ユーザーが自由に編集可能なオンライン百科事典であり、「あらゆる言葉について定義や意味、元ネタを解説する辞書・辞典」をコンセプトとしています。

「単語」だけでなく「句」や「節」や「文」といった一般的な辞典とは離れた内容も編集できるという特徴があり、サブカルチャー文脈の用語(ネットミーム等)が充実しています。

麻雀AI「NAGA」

オンライン麻雀サイト「天鳳」にて十段を達成した麻雀AIのNAGA(Neural Architectural Game Agent)による牌譜解析サービスです。

天鳳の強者の多様な打ち筋をベースに学習したAIで、複数の人間的な強者の打ち筋を学ぶことができます。



研究公開サイト「dmv.nico」

ドワンゴで行っている機械学習系の研究の一部を公開しているサイトです。

