雑学書庫 知識の記録

「ない」は確実、「ある」はたぶん：ブルームフィルタ

記事のURL

https://www.gklibrarykor.com/ja/article/11886/

PDFレイアウトのプレビュー - 本文の一部のみ

アドレスバーに入力したサイトが危険かどうかを、ブラウザが毎回インターネット全体の危険リストをダウンロードして照合するとしたら、遅すぎて重すぎるでしょう。かといって、数百万件のアドレスをすべて端末に持たせるのも負担です。そんなとき、ごくわずかなメモリだけで「これは確実にリストにない」を素早く見分ける道具が使われます。 この道具をブルームフィルタ（Bloom Filter）と呼びます。ある値が集合に入っているかをごく小さな容量で調べる確率的データ構造で、1970年にバートン・ハワード・ブルーム（Burton Howard Bloom）が提案しました。その代わり、「ない」という答えは常に正しく、「ある」という答えはたまに間違うことがあるという独特の約束を抱えています。 この記事では、ブルームフィルタがどんな問題から生まれたのか、値を入れて確認する過程はどう進むのか、誤りの可能性をどう抑えるのか、そして実際のサービスでの使いどころと限界を順に見ていきます。 小さなビット配列で「ない」を見分ける技術 メモリを節約するために、正確さを少しだけ手放す ブルームは、辞書に載った50万語のハイフン処理の問題を例に挙げました。単語の90パーセントは簡単な規則で区切れますが、残りの10パーセントはディスクから別にパターンを読み込む必要がありました。どの単語がその例外なのかを正確に覚えておくには、多くのメ