

Về bài toán nhận diện tên thực thể Y Sinh (Name Entity Recognition - NER) Nhận diện tên thực thể là bài toán xác định tên của các thực thể trong một đoạn văn bản, sau đó phân loại các thực thể đó vào các nhóm đã định nghĩa từ trước. Trong các văn bản thuộc về các miền dữ liệu tổng quan, tên thực thể thường là tên của địa chỉ, tên của tổ chức, tên người. Ví dụ trong văn bản "Tôi sống ở Hà Nội" xuất hiện "Hà Nội" là tên của một thực thể về địa điểm mà thuật toán phải tìm ra. Tuy nhiên, trong văn bản Y Sinh, tên của các thực thể thường là RNA, DNA, các protein, tên các loại gene... Kết quả của việc nhận diện tên thực thể đóng góp một phần không nhỏ để giải quyết những bài toán khai phá dữ liệu trong lĩnh vực Y Sinh. Tên của các thực thể có thể được sử dụng để tạo ra các chú thích về ngữ nghĩa của một văn bản, tạo ra từ khóa đặc trưng cho việc tra cứu. Hay việc nhận diện cũng giúp xây dựng bộ dữ liệu về mối quan hệ thực thể bệnh, thuốc hoặc các triệu chứng đi kèm. Ngoài ra, một ứng dụng khác của bài toán nhận diện tên thực thể là hệ thống trả lời câu hỏi. Các tên thực thể vừa được sử dụng để xử lý truy vấn cũng như tìm kiếm câu trả lời. Hoặc với bài toán trả lời câu hỏi, các kĩ thuật nhận diện tên thực thể được sử dụng như một công cụ giúp chọn ra các câu trả lời phù hợp. Tại hội nghị TREC-8 (Text REtrieval Conference) đã chỉ là 80% các truy vấn dùng để đánh giá các hệ thống này là các câu hỏi "who", "where", "when" [Srihari2000] - "ai", "ở đâu", "khi nào" - vốn có thể trả lời với tên người, tổ chức, địa điểm hoặc ngày tháng.

### **Thêm hình số lượng tăng lên của các bài bài Tin Sinh mỗi năm**

Ngày nay, sự tăng lên của các văn bản trong lĩnh vực Y Sinh càng khiến cho việc khai phá dữ liệu trong lĩnh vực này càng trở nên cần thiết, mà việc phát hiện tên thực thể là một trong đầu vào quan trọng để thực hiện các bài toán khác.

Việc nhận diện tên thực thể Y Sinh ngoài