Trong lĩnh vực lập trình và xử lý chuỗi, thuật toán tìm kiếm chuỗi đóng vai trò quan trọng để tối ưu hóa hiệu suất. Một trong những thuật toán nổi bật là KMP, viết tắt của Knuth-Morris-Pratt, được phát triển vào những năm 1970. Thuật toán này giúp tránh việc so sánh lại các ký tự đã khớp trước đó, nhờ vào việc xây dựng một bảng "prefix" để xác định vị trí nhảy khi có mismatch. Ví dụ, khi tìm kiếm một mẫu trong văn bản dài, KMP có độ phức tạp thời gian tuyến tính O(n + m), với n là độ dài văn bản và m là độ dài mẫu. Thuật toán KMP dùng tìm kiếm chuỗi đặc biệt hiệu quả với văn bản tiếng Việt, nơi có nhiều ký tự có dấu như â, ê, ô, ơ, dẫn đến nhu cầu hỗ trợ bảng chữ cái mở rộng như VietnameseAlphabet. Ngoài ra, KMP được áp dụng rộng rãi trong các ứng dụng như trình soạn thảo văn bản, công cụ tìm kiếm, và phân tích dữ liệu sinh học. Để triển khai, bạn cần xây dựng DFA (Deterministic Finite Automaton) từ mẫu, sau đó quét qua văn bản để tìm vị trí khớp. Trong Java, lớp KMP có thể được kế thừa để hỗ trợ file docx, như trong KMPFile, giúp đọc và tìm kiếm tự động. Thuật toán này vẫn là nền tảng cho nhiều cải tiến hiện đại, chứng tỏ giá trị lâu dài trong khoa học máy tính.