

Tài liệu Yêu cầu Sản phẩm (PRD) – Ứng dụng Xử lý Kết quả Khám Bệnh

1. Tầm nhìn Sản phẩm

Ứng dụng web này ra đời với mục tiêu **tự động hóa việc trích xuất và cá nhân hóa các phiếu kết quả xét nghiệm y tế** từ một tài liệu Word lớn duy nhất.

- **Người dùng mục tiêu:** Quản trị viên phòng khám, nhân sự (HR) hoặc quản lý hồ sơ y tế, những người cần phân phối hiệu quả các phiếu kết quả xét nghiệm riêng lẻ.
- **Giá trị cốt lõi:** Giảm đáng kể công sức thủ công và lỗi trong việc tách và đặt tên các tệp kết quả riêng lẻ, đảm bảo tính toàn vẹn của dữ liệu và phân phối nhanh chóng.

2. Các Tính năng Cốt lõi

2.1. Tải tệp lên (File Upload)

- **Yêu cầu:** Cho phép người dùng tải lên một tài liệu Word duy nhất (định dạng .docx).
- **Tiêu chí chấp nhận:**
 - Hỗ trợ các tệp lớn (3000-5000 trang, dung lượng có thể lên đến hàng trăm MB).
 - Giao diện tải lên thân thiện với người dùng.
 - Có chỉ báo tiến độ (progress indicator) cho việc tải lên các tệp lớn.

2.2. Xử lý & Tách dữ liệu (Data Processing & Separation)

- **Yêu cầu:** Tự động xác định và tách từng phiếu kết quả xét nghiệm riêng lẻ trong tệp Word đã tải lên.
- **Tiêu chí chấp nhận:**
 - Mỗi phiếu được coi là một "trang" logic/đơn vị kết quả riêng biệt trong tài liệu Word gốc, được xác định bởi **mẫu dữ liệu lặp lại** (ví dụ: sự xuất hiện của "Họ và tên:") hoặc **ngắt trang** nếu có.
 - Giữ nguyên toàn bộ nội dung gốc, định dạng (bảng, hình ảnh, kiểu văn bản) và bố cục bên trong mỗi phiếu đã tách.
 - **Trích xuất dữ liệu khóa:** Đối với mỗi trang/phiếu, trích xuất "Họ và tên" (Full Name) và "Code" (Mã số) dựa trên các mẫu cụ thể.
 - **Mẫu ví dụ:** "Họ và tên: Nguyễn Văn A" và "Code: 1001".
 - **Linh hoạt:** AI phải có khả năng xử lý các biến thể nhỏ về khoảng cách hoặc chữ hoa/thường xung quanh "Họ và tên:" và "Code:" (ví dụ: "Họ và tên :", "họ và tên:").

2.3. Tạo tệp (File Generation)

- **Yêu cầu:** Tạo các tệp riêng lẻ (Word hoặc PDF) cho mỗi phiếu kết quả xét nghiệm đã tách.
- **Tiêu chí chấp nhận:**

- **Tùy chọn định dạng đầu ra:** .docx (Word) và .pdf.
- **Quy ước đặt tên:** Mỗi tệp được tạo phải được đặt tên bằng "Code" và "Họ và tên" đã trích xuất theo nguyên tắc: code_ho_va_ten.docx hoặc code_ho_va_ten.pdf.
 - **Ví dụ:** 1001_nguyen_van_a.docx hoặc 1001_nguyen_van_a.pdf.
- **Chuẩn hóa:** Tên đầy đủ phải được chuẩn hóa (chữ thường, không dấu, khoảng trắng thay bằng dấu gạch dưới). Các ký tự đặc biệt khác (ví dụ: dấu gạch nối, dấu nháy đơn, ký hiệu) sẽ bị loại bỏ.

2.4. Tải tệp xuống (File Download)

- **Yêu cầu:** Cung cấp các tùy chọn để tải xuống các tệp đã tạo.
- **Tiêu chí chấp nhận:**
 - Tải tất cả các tệp đã xử lý thành công dưới dạng một tệp lưu trữ .zip duy nhất.
 - Có tùy chọn tải xuống từng tệp riêng lẻ trực tiếp.
 - Hiển thị rõ ràng trạng thái hoàn tất xử lý và sẵn sàng tải xuống.
 - Nếu có lỗi xử lý, tệp .zip chỉ chứa các phiếu thành công và có **thông báo hoặc nhật ký** chỉ rõ số lượng/lý do các phiếu bị lỗi trên giao diện người dùng.

2.5. Giao diện Người dùng (Admin Dashboard)

- **Yêu cầu:** Một giao diện quản trị đơn giản để quản lý các tệp đã tải lên và kết quả đã tạo.
- **Tiêu chí chấp nhận:**
 - Liệt kê các tệp Word lớn đã tải lên.
 - Hiển thị trạng thái xử lý (đang chờ, đang xử lý, đã hoàn tất, thất bại).
 - Truy cập các tùy chọn tải xuống.
 - Xác thực người dùng cơ bản (một người dùng quản trị duy nhất được cấu hình sẵn ban đầu).

3. Ngăn xếp Công nghệ (Technical Stack)

- **Frontend Framework:** React (hoặc một framework hiện đại phù hợp cho các ứng dụng web tương tác).
- **Backend Language/Framework:** Python (Flask/Django) hoặc Node.js (Express) - phù hợp cho việc xử lý tệp và API.
- **Database:** PostgreSQL (hoặc MongoDB cho tính linh hoạt).
- **Thư viện xử lý tệp:**
 - **Python:** python-docx (để đọc và sửa đổi tệp Word), PyPDF2 hoặc ReportLab (để tạo PDF), re (cho biểu thức chính quy để trích xuất dữ liệu).
 - **Node.js:** mammoth (để chuyển đổi docx sang HTML), pdf-lib (để tạo PDF), hoặc các thư viện tương tự khác cho xử lý docx và PDF.
- **Nền tảng triển khai (Deployment Platform):** AWS (EC2/Lambda/S3) hoặc Google Cloud (Compute Engine/Cloud Functions/Cloud Storage) để có khả năng mở rộng và xử lý các tệp lớn.
- **Xác thực:** Đăng nhập bằng tên người dùng/mật khẩu cơ bản (cấu hình sẵn).

4. Thiết kế & Trải nghiệm Người dùng (Design & User

Experience)

- **Thiết kế hình ảnh:** Giao diện chuyên nghiệp, tối giản, sạch sẽ. Tập trung vào tính dễ sử dụng cho người dùng không chuyên về kỹ thuật.
- **Thiết kế đáp ứng (Responsive Design):** Tối ưu hóa chủ yếu cho việc sử dụng trên máy tính để bàn, nhưng vẫn hoạt động tốt trên máy tính bảng.
- **Xử lý lỗi (Error Handling):** Thông báo lỗi rõ ràng, thân thiện với người dùng cho các trường hợp tải lên thất bại, sự cố xử lý hoặc lỗi trích xuất dữ liệu.

5. Tiêu chuẩn Hiệu suất & Chất lượng (Performance & Quality Standards)

- **Tốc độ xử lý:** Hướng tới việc xử lý hiệu quả; cung cấp ước tính cho các kích thước tệp khác nhau nếu có thể. (ví dụ: "Xử lý 3000 trang trong vòng X phút").
- **Độ chính xác:** Độ chính xác cao trong việc trích xuất dữ liệu và giữ nguyên nội dung.