



# HƯỚNG DẪN IMPORT HÀNG LOẠT BÀI BÁO

## 🎯 Mục đích

Công cụ này giúp bạn import hàng loạt **1200+ bài báo cũ** vào hệ thống, bao gồm:

- Bài báo đã xuất bản (PUBLISHED)
- Bài báo không được duyệt (REJECTED)
- Upload PDF files lên S3
- Tự động tạo tác giả, danh mục, số tạp chí
- Phân quyền xem: Bài REJECTED chỉ admin/editor/tác giả xem được



## CÁC BƯỚC THỰC HIỆN

### Bước 1: Chuẩn bị dữ liệu

#### 1.1. Tải template Excel

Tải file mẫu: [IMPORT\\_TEMPLATE.xlsx](#) (./IMPORT\_TEMPLATE.xlsx)

#### 1.2. Điền thông tin vào Excel

CÁC CỘT TRONG TEMPLATE:

Cột	Tên cột	Bắt buộc	Mô tả	Ví dụ
A	<b>STT</b>	✓	Số thứ tự	1, 2, 3...
B	<b>Mã bài báo</b>	✓	Mã định danh duy nhất	BB-2020-001
C	<b>Tiêu đề (VN)</b>	✓	Tiêu đề tiếng Việt	Ứng dụng AI trong y tế
D	<b>Tiêu đề (EN)</b>	✗	Tiêu đề tiếng Anh	AI Applications in Healthcare
E	<b>Tác giả</b>	✓	Họ tên tác giả	Nguyễn Văn A
F	<b>Email tác giả</b>	✓	Email tác giả	nguyen-vana@example.com
G	<b>Đơn vị</b>	✓	Nơi công tác	Đại học Quốc gia Hà Nội
H	<b>Tóm tắt (VN)</b>	✓	Tóm tắt tiếng Việt	Nghiên cứu này tập trung vào...
I	<b>Tóm tắt (EN)</b>	✗	Tóm tắt tiếng Anh	This research focuses on...
J	<b>Từ khóa</b>	✓	Từ khóa, phân cách bởi dấu phẩy	AI, Machine Learning, Healthcare
K	<b>Danh mục</b>	✓	Tên danh mục	Công nghệ thông tin
L	<b>Năm xuất bản</b>	✓	Năm (YYYY)	2020
M	<b>Số tạp chí</b>	✓	Số thứ tự của tạp chí	1, 2, 3...
N	<b>Tập tạp chí</b>	✗	Tập/Volume	15
O	<b>Trang bắt đầu</b>	✗	Trang đầu	1
P	<b>Trang kết thúc</b>	✗	Trang cuối	10
Q	<b>Trang số</b>	✗	Định dạng trang	1-10 hoặc 5-15
R	<b>Trang số format</b>	✗	Định dạng khác	pp. 1-10

Cột	Tên cột	Bắt buộc	Mô tả	Ví dụ
S	<b>Tên file PDF</b>	✓	Tên file PDF (chính xác)	article-001.pdf
T	<b>Trạng thái</b>	✓	PUBLISHED hoặc REJECTED	PUBLISHED
U	<b>DOI</b>	✗	DOI (nếu có)	10.1234/example.2020.001
V	<b>Ghi chú</b>	✗	Ghi chú bổ sung	Bài viết hay, cần review

#### LƯU Ý QUAN TRỌNG:

- ✓ **Tên file PDF** phải khớp chính xác với file trong folder `pdf-imports/`
- ✓ **Email tác giả** phải là email hợp lệ
- ✓ **Trạng thái**:
  - `PUBLISHED` = Bài đã đăng → Hiển thị công khai
  - `REJECTED` = Bài không duyệt → Chỉ admin/editor/tác giả xem được

## Bước 2: Chuẩn bị file PDF

### 2.1. Tạo folder `pdf-imports/`

```
cd /home/ubuntu/tapchi-hcqs/nextjs_space/scripts
mkdir -p pdf-imports
```

### 2.2. Copy tất cả file PDF vào folder này

```
# Ví dụ: Copy từ folder khác
cp /path/to/your/pdfs/*.pdf ./pdf-imports/
```

#### Cấu trúc thư mục:

```
scripts/
  └── import-articles-from-excel.ts
  └── IMPORT_TEMPLATE.xlsx
  └── IMPORT_GUIDE.md
  └── articles-import.xlsx          - File Excel của bạn
  └── pdf-imports/                  - Folder chứa PDF
    └── article-001.pdf
    └── article-002.pdf
    └── article-003.pdf
    └── ...
```

## Bước 3: Chạy script import

### 3.1. Cài đặt dependencies (nếu chưa có)

```
cd /home/ubuntu/tapchi-hcqs/nextjs_space
yarn add exceljs
```

### 3.2. Chạy script

```
cd /home/ubuntu/tapchi-hcqs/nextjs_space
yarn tsx scripts/import-articles-from-excel.ts scripts/articles-import.xlsx
```

#### Output mẫu:

```
=====
[!] BẮT ĐẦU IMPORT BÀI BÁO TỪ EXCEL + PDF
=====

[?] Đọc file Excel: scripts/articles-import.xlsx
[✓] Đã đọc 1200 dòng dữ liệu hợp lệ

[?] Tổng số bài báo cần import: 1200
-----

[1] Import: Ứng dụng AI trong y tế
[?] Tạo tác giả mới: Nguyễn Văn A (nguyenvana@example.com)
[?] Tạo danh mục mới: Công nghệ thông tin
[?] Tạo Volume mới: Tập 15, Năm 2020
[?] Tạo Issue mới: Số 1, Tập 15, Năm 2020
[?] Upload PDF: article-001.pdf (2.5 MB)
[✓] Upload thành công: articles/1733805123456-BB-2020-001-article-001.pdf
[✓] Tạo Submission: BB-2020-001
[✓] Tạo Article: ID 1
[!] Hoàn tất import bài báo #1

[2] Import: Machine Learning trong giáo dục
...
-----

[!] KẾT QUẢ IMPORT
=====

[✓] Thành công: 1180/1200
[✗] Thất bại: 20/1200

[!] CHI TIẾT LỖI:
[Dòng 15] File PDF không tồn tại: missing-file.pdf
[Dòng 89] Email không hợp lệ: invalid-email
-----
```

# KIỂM TRA KẾT QUẢ

## 1. Kiểm tra trong Database

```
cd /home/ubuntu/tapchi-hcqs/nextjs_space
yarn tsx << 'EOF'
import { PrismaClient } from '@prisma/client';
const prisma = new PrismaClient();

async function check() {
  const published = await prisma.submission.count({ where: { status: 'PUBLISHED' } });
  const rejected = await prisma.submission.count({ where: { status: 'REJECTED' } });
  const articles = await prisma.article.count();
  const authors = await prisma.user.count({ where: { role: 'AUTHOR' } });

  console.log(`\n${colorfulIcon} THÔNG KÊ IMPORT:`);
  console.log(`    Bài đã xuất bản (PUBLISHED): ${published}`);
  console.log(`    Bài Không duyệt (REJECTED): ${rejected}`);
  console.log(`    Tổng Article records: ${articles}`);
  console.log(`    Tổng tác giả: ${authors}\n`);
}

check().finally(() => prisma.$disconnect());
EOF
```

## 2. Xem trên giao diện

### Bài đã xuất bản (PUBLISHED):

- ✓ Hiển thị trên trang chủ
- ✓ Hiển thị trong /archive
- ✓ Hiển thị trong /articles
- ✓ Mọi người đều xem được

### Bài không duyệt (REJECTED):

- ✗ KHÔNG hiển thị công khai
- ✓ Admin/Editor xem được trong dashboard
- ✓ Tác giả xem được bài của mình
- ✓ Có thể tra cứu nếu biết mã bài

## ⚠ XỬ LÝ LỖI THƯỜNG GẶP

### Lỗi 1: File PDF không tồn tại

⚠ File PDF không tồn tại: article-123.pdf

**Nguyên nhân:** Tên file trong Excel không khớp với file thực tế.

#### Giải pháp:

- Kiểm tra tên file PDF trong folder pdf-imports/
- Sửa lại tên file trong Excel cho khớp (phân biệt HOA/thường)
- Chạy lại script

## Lỗi 2: Email không hợp lệ

 Lỗi: Email không hợp lệ

**Nguyên nhân:** Email tác giả bị sai format hoặc trống.

**Giải pháp:**

1. Kiểm tra email trong Excel
2. Đảm bảo format đúng: `user@domain.com`
3. Chạy lại script

## Lỗi 3: Upload S3 thất bại

 Lỗi upload PDF: Access Denied

**Nguyên nhân:** Thiếu cấu hình AWS S3.

**Giải pháp:**

1. Kiểm tra file `.env`:

```
bash
AWS_ACCESS_KEY_ID=your-access-key
AWS_SECRET_ACCESS_KEY=your-secret-key
AWS_REGION=ap-southeast-1
AWS_BUCKET_NAME=your-bucket-name
```

2. Test S3 connection:

```
bash
aws s3 ls s3://your-bucket-name
```



## BẢO MẬT & PHÂN QUYỀN

### Bài PUBLISHED (Đã xuất bản)

-  **Public** - Mọi người xem được
-  Hiển thị trong search
-  Có thể download PDF
-  Hiển thị trong Archive

### Bài REJECTED (Không duyệt)

-  **Private** - Chỉ những người sau xem được:
  1. **SYSADMIN** - Quản trị hệ thống
  2. **EIC** - Tổng biên tập
  3. **MANAGING\_EDITOR** - Biên tập điều hành
  4. **SECTION\_EDITOR** - Biên tập chuyên mục
  5. **TÁC GIẢ** - Chỉ xem được bài của chính họ
-  Không hiển thị trong search công khai

- ✗ Không hiển thị trong Archive
- ✓ Có thể tra cứu trong Admin Dashboard

## THỐNG KÊ SAU KHI IMPORT

### Truy vấn thống kê:

```
-- Tổng số bài đã xuất bản
SELECT COUNT(*) FROM "Submission" WHERE status = 'PUBLISHED';

-- Tổng số bài không duyệt
SELECT COUNT(*) FROM "Submission" WHERE status = 'REJECTED';

-- Thống kê theo danh mục
SELECT
    c.name AS category,
    COUNT(*) AS total,
    SUM(CASE WHEN s.status = 'PUBLISHED' THEN 1 ELSE 0 END) AS published,
    SUM(CASE WHEN s.status = 'REJECTED' THEN 1 ELSE 0 END) AS rejected
FROM "Submission" s
JOIN "Category" c ON s."categoryId" = c.id
GROUP BY c.name
ORDER BY total DESC;

-- Thống kê theo năm
SELECT
    i.year,
    i.number AS issue_number,
    COUNT(a.id) AS articles_count
FROM "Issue" i
LEFT JOIN "Article" a ON a."issueId" = i.id
GROUP BY i.year, i.number
ORDER BY i.year DESC, i.number DESC;
```

## MẸO HỮU ÍCH

### 1. Test với số lượng nhỏ trước

Trước khi import 1200 bài, hãy test với 10-20 bài:

1. Tạo file Excel với 10-20 dòng đầu
2. Copy 10-20 PDF tương ứng vào `pdf-imports/`
3. Chạy script
4. Kiểm tra kết quả
5. Nếu OK, import toàn bộ

### 2. Backup database trước khi import

```
cd /home/ubuntu/tapchi-hcqs/nextjs_space
bash scripts/backup-db.sh
```

### 3. Đặt tên file PDF có quy tắc

Ví dụ:

- 2020-01-001.pdf (Năm-Số-STT)
- BB-2020-001.pdf (Prefix-Năm-STT)
- AI-Healthcare-2020.pdf (Chủ đề-Năm)

### 4. Sử dụng Excel Formula để tự động

Tự động tạo mã bài báo:

```
= "BB- " & L2 & "-" & TEXT(A2, "000")
// Kết quả: BB-2020-001
```

Tự động tạo tên file PDF:

```
= B2 & ".pdf"
// Kết quả: BB-2020-001.pdf
```



Nếu gặp vấn đề:

1.  Đọc kỹ log lỗi
2.  Kiểm tra file Excel và PDF
3.  Kiểm tra cấu hình AWS S3
4.  Backup database trước khi thử lại



### CHECKLIST TRƯỚC KHI IMPORT

- [ ]  Đã tải template Excel
- [ ]  Đã điền đầy đủ thông tin vào Excel
- [ ]  Đã tạo folder pdf-imports/
- [ ]  Đã copy tất cả PDF vào folder
- [ ]  Tên file PDF khớp với Excel
- [ ]  Đã cài exceljs : yarn add exceljs
- [ ]  Đã cấu hình AWS S3 trong .env
- [ ]  Đã backup database
- [ ]  Đã test với 10-20 bài trước
- [ ]  Sẵn sàng import toàn bộ!

Chúc bạn import thành công! 🎉