IANEC：当代作家数字档案调查

GREYC 实验室开展的工作

Emmanuel GIGUET

卡昂诺曼底大学、ENSICAEN、法国国家科学研究中心、诺曼底大学联盟、
GREYC第6072联合研究单位，F-14000 卡昂，法国 emmanuel.giguet@cnrs.fr

2025年11月

目录

IANEC 项目.1
原生数字语料库的技术分析相关工作.3
1. 借助哈希数据库分离系统文件和应用程序与用户文件.3
2. 读取 Mac 分区中的文件并创建 FUSE 模块.6
3. 操作系统虚拟化与仿真界面.7
4. 改进 G’DIP 界面.9
数字文档语义分析相关工作.10
5. 文本文件自动分类.10
6. 图像分析与分类（文档、照片、标志）.12
7. 仇恨言论检测与档案可传播性.14
总结.16
GREYC 指导实习.17
8. 2024年指导的实习.17
9. 2025指导的实习.17

由文化部支持的项目

MINISTÈRE
DE LA CULTURE
Liberté
Égalité
Fraternité

IANEC 项目

IANEC 项目，即当代作家数字档案调查，由卡昂计算机、图像与仪器研究组（GREYC）牵头，并由文化部通过创新数字服务（SNI）项目征集计划资助，旨在开发专用数字取证工具，以自动分析当代出版记忆研究所（IMEC）的档案语料库。该项目基于这样一个事实：在当代档案中心，数字档案的接收日益频繁，而数字技术对于原生数字语料库的提取、识别、处理和描述也变得不可或缺。

该项目分两个阶段开展：

- • 阶段 1：2024年，项目第一阶段的工作包括建立数据，开发 GREYC 和 IMEC 方面的软件工具，并利用 IANEC 项目资金招募学生参与项目和实习。GREYC 的三名实习生在这一阶段完成了一些开发工作。其中部分基于 G’DIP 平台实现的开发成果于2024年由 IMEC 进行了测试。

- 阶段 2：2025年上半年，GREYC 招募了 6 名实习生。上半年期间持续开展开发工作，IMEC 方面则安排了一名实习生跟进并测试开发的工具。这些实习生的经费均由该项目提供。2025年开发的工具和功能有意设计为独立于 G’DIP 平台，以便促进互操作性，并加快项目中的开发和测试阶段。

IANEC 项目旨在从两个发展方向满足 IMEC 的需求：

1. “分析原生数字语料库的技术特征”方向，目标是针对给定的数字档案语料库生成一份详尽的技术报告，并应用和扩展由 GREYC 开发的 G'DIP 数字取证平台的功能。2. “分析文档内容”方向，旨在识别并自动生成有助于档案描述的特定信息。

这项工作由 GREYC 实验室的 Christophe Rosenberger（ENSICAEN 大学教授）、Emmanuel Giguet（法国国家科学研究中心研究员）和 Tanguy Gernot（法国国家科学研究中心研究工程师）共同协调。三人均为 SAFE 网络安全研究团队数字取证主题的成员。相关内容将在以下两个章节中介绍，随后进行总结。

与原生数字语料库技术分析相关的工作

1. 借助哈希数据库，将系统文件和应用程序与用户文件区分开来

在磁盘映像中，系统文件和应用程序产生的文件占据了数据总量的很大一部分。它们数量众多，增加了档案工作者的工作难度；档案工作者必须花费大量时间识别并排除这些文件，才能专注于用户实际创建的文件。

在 IANEC 项目中，我们的目标是自动识别磁盘映像中的技术文件，以便将系统文件和应用程序产生的文件与用户实际创建的文件区分开来。这种区分对 IMEC 的档案工作者以及更广泛地说，对所有开展数字取证的调查人员都很重要：它可以快速筛选出相关文件，让专家将注意力集中在有价值的个人文件上，例如信件、合同文件、发票、照片和电子邮件。

为实现这一目标，我们采用的方法是参照美国国家标准与技术研究院（NIST）的参考数据集（RDS）构建哈希数据库，并根据项目需求进行调整。主要问题在于，NIST 的官方数据库未涵盖 Macintosh 系统和应用程序的旧版本，而这些版本对于档案工作者，尤其是 IMEC 的档案工作者所处理的馆藏至关重要。因此，我们必须针对经典 Mac 环境开展专门工作，以补充和丰富这些数据库。我们通过仿真不同版本的 macOS，并在这些环境中安装应用程序，提取操作系统和应用程序文件的特征码。

具体而言，开发了专用工具，用于提取磁盘映像或文件夹的特征码，并自动将其录入哈希数据库。我们特别注重管理操作系统、应用程序和文件之间的关联，以避免不必要地重复数据，并确保系统自带的文件不会与应用程序添加的文件混淆。该数据库采用 PostgreSQL 实现；与简单的 SQLite 文件相比，PostgreSQL 更适合此用途，因为它支持网络访问、索引和协作管理。

我们开发了图形工具和命令行工具 `hash_match_manager` ，方便档案工作者与数据库交互，并在需要时添加新的特征码。我们还为 Autopsy 数字取证平台开发了一个插件，以便在取证环境中集成这一按重要性筛选文件的功能。此外，我们并行开发了一款独立工具，可生成交互式分析报告。

最终，该项目建立了一个可用且可扩展的特征码数据库，用于区分系统文件和应用程序生成的文件，以及用户实际创建的文件

。项目还设计了用于自动添加新特征码的工具。两个界面可在 IMEC 提供的磁盘映像上，具体实现识别用户文件这一功能。尽管仍存在一些局限，尤其是对 Mac 系统的覆盖有限，但这项工作让特征码数据库得以实际应用和协作使用，并展现出大幅节省档案工作时间的潜力，也将惠及更广泛的数字取证领域。

参考文献：

Titouan LE BRET - 4个月。当代作家数字档案调查：数字取证分析与数字档案处理。由 Tanguy Gernot、Emmanuel Giguet 和 Christophe Rosenberger 指导的硕士一年级实习。2025年3月3日至7月4日。[实习报告]

Guillaume HAUTOT - 4个月。当代作家数字档案调查。由 Tanguy Gernot、Emmanuel Giguet 和 Christophe Rosenberger 指导的硕士一年级实习。2025年3月3日至7月4日。[实习报告]

示意图：

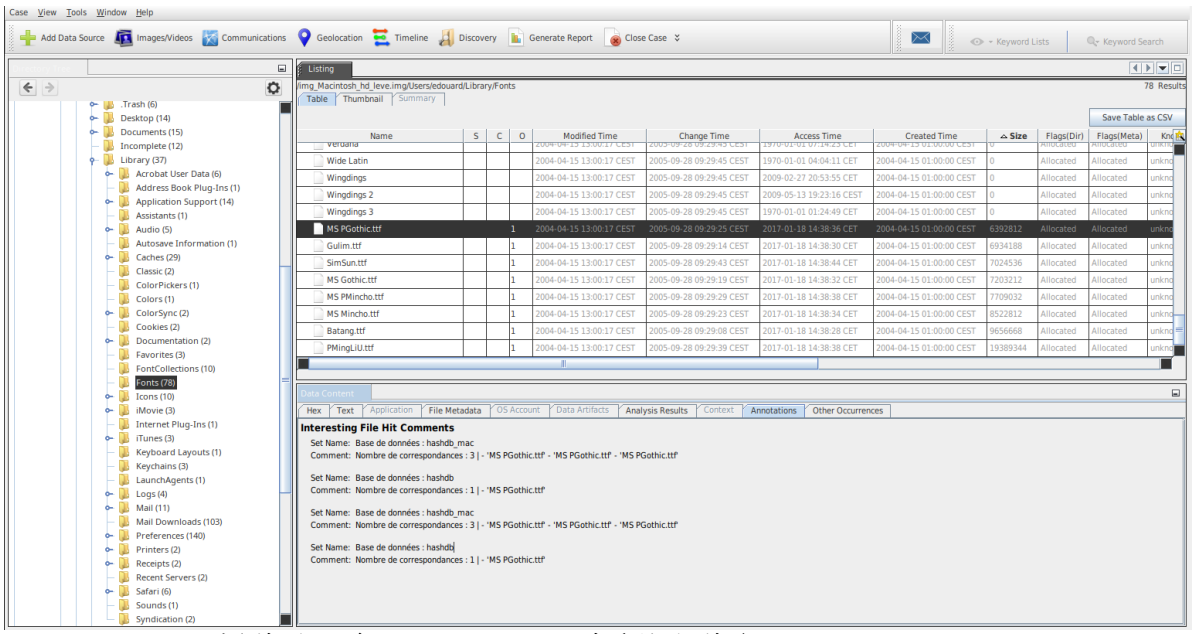


图*1*：*Autopsy* 插件的示意图，可用于识别系统文件和应用程序文件



Rapport d'analyse de la source Macintosh_hd_leve.img

Pour cette analyse, seuls ont été conservés les fichiers qui ne sont pas de taille nulle.

Informations concernant l'analyse :

- Date : 6 juin 2025 à 10h59:40 (Heure française)
- Informations sur la/les base(s) de données :
  - Base de données : **hashdb** | version : 2025.03.20 - GREYC-legacy-minimal
  - Base de données : **rds_legacy_minimal** | version : 2025.03.1 - legacy
  - Base de données : **hashdb_mac** | version : 2025.04.25 - GREYC-mac_os_6_to_10
- Nom de la data source : Macintosh_hd_leve.img
- Nom de l'examinateur : root

Liste des fichiers sans correspondance :

10 Changer

Nom du fichier	Taille	Hash MD5
dmg/Desktop%20DF	79554	9832c6dfbd563bbf6066a5b2dac0da36
dmg/.fseventsd/000000000001e0cb	70	7f3126d778be0d2ae446f0cea4d58d3d
dmg/Applications/ABBYY%20FineReader%205%20Sprint/Readme	2392	a6cb086c916d7427a13b0963cfdb6d1d
dmg/Desktop%20DB	37376	560aefecb7ffd63f1cd4010335dab593
dmg/.fseventsd/000000000001e0ce	38	ebb12e59f60c213f883d07e17ff3f2d5
dmg/Applications/ABBYY%20FineReader%205%20Sprint/License	9515	a40304a939f1564f5f30c400f49c2135
dmg/.fseventsd/000000000001e0cd	71	c266b7935d4cbdf591c7ae720a965e7f
dmg/.hotfiles.btree	196608	dd646dcc46b5a6d84713c9f7e6f08e0d
dmg/.DS_Store	12292	613e303488a4f22167dc2075ead576eb
dmg/Applications/ABBYY%20FineReader%205%20Sprint/Launch%20FineReader%205%20Sprint	175489	2fbc46371ee6550e8e5ab728ff335a8e

« Précédent Page 1 / 17898 Suivant » Lien vers JSON (117834 Ko)

OS détectés par des correspondances sur la source:

Nom de l'OS	Version	# occurrences	Base de données
Mac OS X Tiger	10.4	131832	hashdb
Mac OS 7	7.6.1	42	hashdb
Mac OS X Tiger	10.4	133655	hashdb_mac
Mac OS X Panther	10.3	55217	hashdb_mac
Mac OS X Jaguar	10.2	24806	hashdb_mac
Mac OS 9	9.0.5	1073	hashdb_mac
Mac OS 8	8.6	15	hashdb_mac
Mac OS 8	8.5	15	hashdb_mac
MacOS 8	8.0	15	hashdb_mac
Mac OS 7	7.1 (m68k)	10	hashdb_mac

Correspondances avec des packages connus :

Correspondances détectées dans la base de données : hashdb

ID	Nom du package	Version	Langue	OS	% de correspondance	# occurrences	Détails
2	Mac OS X Tiger(vierge)	10.4	French	Mac OS X Tiger	78%	126454	Afficher/Masquer les fichiers
5	Adobe Acrobat	6.0	French	Mac OS X Tiger	138%	4142	Afficher/Masquer les fichiers
4	StuffitDeluxe	11	French	Mac OS X Tiger	11%	188	Afficher/Masquer les fichiers
3	Microsoft Office 2004	10.4	French	Mac OS X Tiger	63%	1048	Afficher/Masquer les fichiers

图2：关于磁盘映像中系统文件和应用程序文件的 *HTML* 格式分析报告示意图

2. 读取 Mac 分区上的文件并创建 FUSE模块

访问 macOS 使用的某些类型分区中的数据往往很复杂，或者至少远不如 FAT、NTFS 或 EXT 等更常见的分区格式简单。这一难题尤其涉及较早期的 Macintosh 文件系统（MFS、HFS、HFS+），这些系统保留了一些特殊机制，例如对“资源分叉”的管理；资源分叉中包含数字取证所需的重要元数据。

对于档案工作者而言，这种情况使得直接利用磁盘映像变得困难，因此不得不借助专用仿真器，或采取繁琐的操作，例如将数据复制到便于取证的分区类型中。因此，这项工作的目标是提供一种解决方案，使这些分区能够在 Linux 下轻松读取，并让其中的所有信息（包括资源分叉）都能直接访问，从而简化并自动化对映像的批量处理。

为此，我们选择开发一个基于 FUSE 的模块，使磁盘映像能够在用户空间中挂载，并像其他普通文件系统一样被读取。该实现基于 libhfs库，以 C语言完成。该模块会展示带有数据分叉和资源分叉的文件，后者出现在与常规结构平行的虚拟目录 `.rsrc` 中，从而保留原始文件系统的逻辑。

开发过程中需要实现一组 FUSE函数，用于处理目录树中的打开、读取和导航。通过单元测试、使用 CMake 以及 Valgrind 等内存检查工具，确保代码的稳健性。该模块以只读方式运行，能够高效访问 Mac 磁盘映像的内容，但某些功能仍受到限制，例如 Mac 特有的符号链接管理以及多线程支持。尽管仍有改进之处，这个名为 `hfuse`的模块已成为档案工作者以及更广泛的调查专业人员可操作且实用的工具，他们现在可以直接在 Linux 环境中浏览和利用 HFS/HFS+ 磁盘映像。

参考：

Guillaume HAUTOT - 4个月。当代作家数字档案调查。由 Tanguy Gernot、Emmanuel GIGUET 和 Christophe Rosenberger 指导的硕士一年级实习。2025年3月3日至7月4日。[实习报告]

3. 操作系统虚拟化与仿真界面

对 IMEC 数字档案的利用面临诸多技术难题，这些问题与原始环境的过时密切相关。部分文档来自20世纪90年代至2010年的 Macintosh 系统。这些环境带来了若干问题：首先是硬件过时，因为 Mac OS Classic 运行于如今已不复存在的 PowerPC 或 M68k 硬件架构之上。其次是软件过时，许多文件格式（如 ClarisWorks、QuarkXPress）已无法通过现有软件直接读取。最后，互操作性方面仍存在诸多限制：即使文件能够读取，也必须确保其在现代环境中的完整性和一致性，同时不丢失元数据。

为应对这些问题，该项目设定了双重目标。首要目标是构建仿真环境，无论 Mac OS 的哪个版本，都能对其进行忠实复现，从而使用户能够在文档原始的软件环境中访问这些文档。第二个目标是通过设计一款配备统一界面的软件工具，简化这些环境的使用，并屏蔽不同仿真系统所带来的复杂性，使档案工作者无需直接操作复杂的仿真器，也能利用这些技术。

这种方法对数字保存意义重大。仿真能够确保用户在接近文档生成时的条件下访问文档，从而保存其结构、外观，有时甚至保留其使用体验。通过规范化并自动执行虚拟磁盘准备、机器配置等任务，所开发的界面降低了出错风险，也让非专业用户能够轻松操作。因此，该工具回应了双重需求：忠实保存数字遗产，并使其日常利用更加便捷。

为实现这些成果，使用了两项主要技术。SheepShaver 能够高效仿真 Mac OS Classic 7.5.2 至 9.0.2 版本，性能良好，且拥有活跃的社区。QEMU 的配置较为复杂，但可用于仿真 PowerMac G3/G4 环境，并运行早期版本的 Mac OS X（最高至 10.5）。这两款工具相辅相成，满足了所有需求。与此同时，我们使用 Rust 和 Ratatui 库开发了一个文本界面。该界面将不同的仿真器整合在同一个应用程序中，自动管理配置文件和磁盘副本，并采用模块化架构，便于添加新的环境。

所取得的成果具体且可直接投入使用。现在只需几个步骤，即可启动预先配置好的 Mac OS Classic 或 Mac OS X 环境，加载来自档案的磁盘，并在不更改原始数据的情况下浏览其内容。该界面支持在副本上操作、连接新的存储介质以及复制现有环境，同时确保与 IMEC 的基础设施兼容。尽管仍有一些局限（QEMU 的性能有待提升，且某些原始磁盘无法直接启动），该项目最终开发出了一款实用工具，大大简化了档案环境中的仿真操作。

参考文献：

Hamid BEN OMAR——4个月。将 G’DIP 取证模块集成到 Autopsy，并为对旧版 Mac OS 进行数字取证做好准备。由 Emmanuel Giguet 和 Tanguy Gernot 指导的二年级实习。2024年4月15日至2024年8月19日。[实习报告]

Matthias DAVID - 4个月。当代作家数字档案调查：为当代出版记忆研究所提供创新的数字服务。由 Tanguy Gernot、Emmanuel Giguet 和 Christophe Rosenberger 指导的硕士一年级实习。2025年3月3日至7月4日。[实习报告]

示意图：

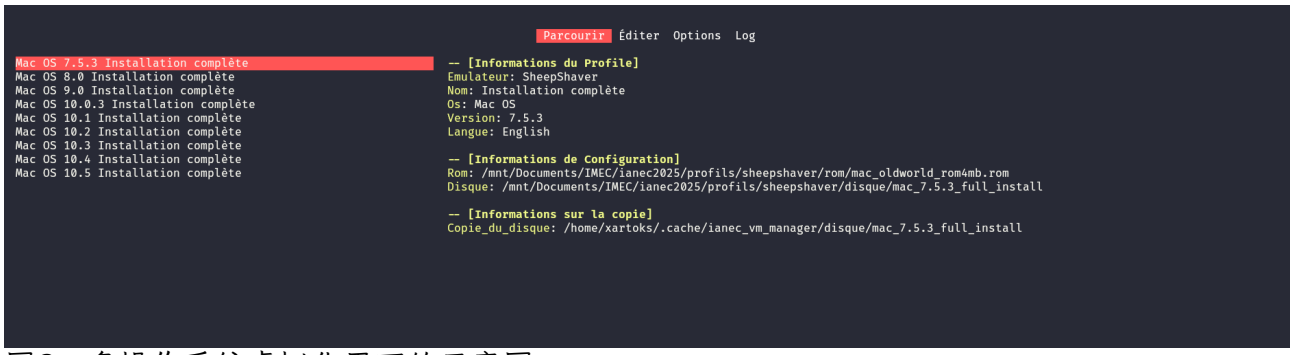


图*3*：多操作系统虚拟化界面的示意图

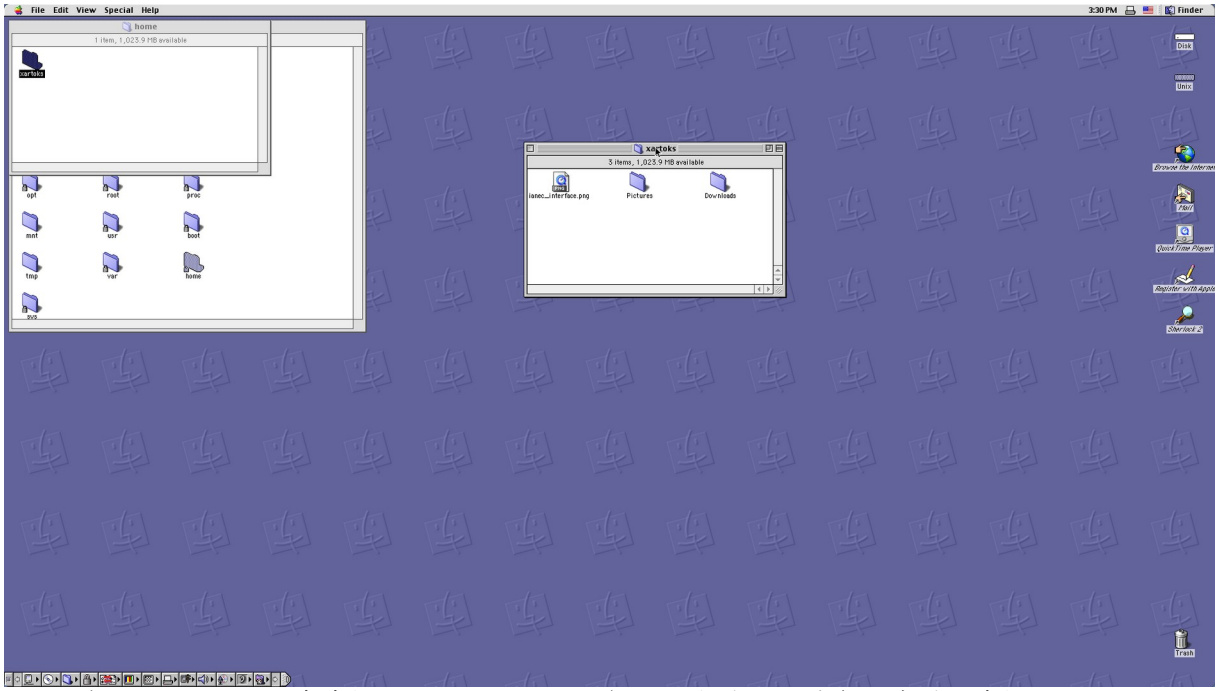


图*4*：在 *SheepShaver* 中虚拟化 *MacOS 9* 的示意图，该系统通过多操作系统虚拟化界面启动

4. G’DIP 界面的改进

由于在 GREYC 内研发的 G’DIP 数字取证平台从未在实验室之外使用过，因此与 IMEC 的合作凸显了重新设计界面人机工程的必要性，以便用户更容易上手。本次实习的目的正是如此。

参考文献：

Arthur LOMBARDO - 2个月。改进 G’DIP 数字取证平台的用户界面。由 Emmanuel Giguet 和 Tanguy Gernot 指导的硕士一年级实习。时间为2024年5月27日至7月19日。

与数字文档语义分析相关的研究工作

5. 文本文件自动分类

IMEC 的数字档案馆藏包含大量类型各异的文本文件：私人信件、文学草稿、私人日记、发票以及合同。数据种类繁多，加之档案全宗规模庞大，使得人工分析耗时、繁琐，也难以扩展到所有待处理数据。

文本文件自动分类研究旨在建立分类器，根据适用于归档的类别（如行政事务、通信、写作活动、私人生活等），实现对这些文档的自动整理。IMEC以分类体系的形式提出了这些类别。自动分类系统有助于档案工作者根据个人文件的类型进行整理，从而间接缩短每个档案全宗的处理时间。

L项目采用的方法主要基于两大方向 :

- 传统机器学习：利用结合词汇类别的词语向量表示，训练传统监督模型，例如支持向量机、随机森林、XGBoost。

- 迁移学习和大型语言模型：采用 BART 或 Mistral 等模型，并针对档案类别进行适配。开展了多项实验 s
  例如对 BART 模型进行专门化，或使用 Mistral 探索摘要生成或自动描述。

交叉比较表明，两种方法互为补充：机器学习在处理明确的词汇结构时非常可靠，而语言模型则更善于捕捉语义细微差别。分歧主要出现在边界案例中，此时仍需人工核验。最后，两项补充实验——对主题相似的文档进行聚类，以及采用检索增强生成（RAG）——展现了结合自动分类与语义检索的价值，并为开发用于探索档案全宗的交互式界面开辟了道路。

参考文献：

Dihia SLIMANA - 4个月。当代作家数字档案调查。由 Tanguy Gernot、Emmanuel Giguet 和 Christophe Rosenberger 指导的硕士一年级实习。2025年3月3日至7月4日。 [实习报告]

Mohammed Salah AISSAOUI - 4个月。当代作家数字档案调查。由 Tanguy Gernot、Emmanuel GIGUET 和 Christophe Rosenberger 指导的硕士一年级实习。2025年3月3日至7月4日。[实习报告]

示意图：

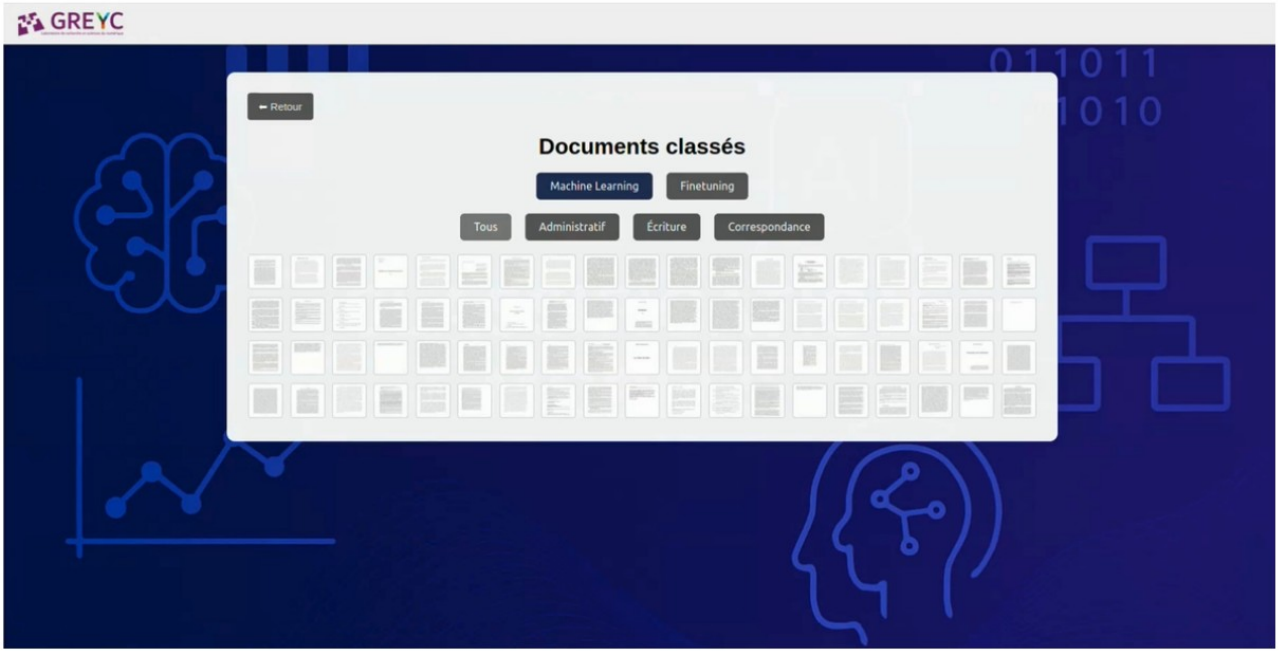


图*5*：原生数字文本分类平台示意图



图*6*：基于本地文档的检索增强生成聊天机器人原型示意图

6. 图像分析与分类（文档、照片、标志）

数字档案馆藏也包含种类繁多的图像：印刷或手写文档的扫描图像、个人照片、屏幕截图、图标和标志。同样，主要难点在于文件类型繁杂，此外图像质量也往往参差不齐。

因此，这项工作的目标有两个：一是区分主要的视觉类别（文档、人物照片、图标/标志、其他图像）；二是提取文档图像中的文本内容，并将其纳入整体分类。开展这项工作的价值在于突出承载信息的文档，同时过滤掉附属或装饰性元素。

测试了两种用于提取文档图像文本内容的主要工具：一种是传统光学字符识别工具 Tesseract，它很快就显现出局限性，无法很好地处理手写、倾斜或质量较差的文本。此外，该工具无法自动区分有价值的图像与不具档案价值的残余图像，因此还需要后续进行更高层次的处理。一种创新工具 Mistral-OCR 的表现远胜于 Tesseract，尤其是在处理手写文本和退化文本方面。

测试了多种互补的分类策略：

- 基于深度学习的视觉分类：通过微调，对 ResNet 等预训练模型，尤其是视觉 Transformer（ViT），进行了适配。所得结果在四分类（照片、文档、图标、其他）方面非常有前景。不过，这些结果仍需在其他语料库上验证。

- 多模态模型：测试了 PixTral-12B，将视觉分析与描述生成相结合，从而提高了结果的可读性，并减少了幻觉导致的错误。

这些实验催生了可投入使用的原型，能够载入完整文件夹，同时对文档和图像进行分类、查看预测结果并导出结果。这些工具由此提供了一套连贯的流程，用于筛选、描述和利用数字档案全宗。

参考文献：

Thomas VARIN - 3个月。档案图像中的人脸检测与识别。由 Emmanuel Giguet 和 Tanguy Gernot 指导的二年级实习。2024年6月3日至8月23日。[实习报告]

Dihia SLIMANA - 4个月。当代作家数字档案调查。文档自动分类。由Tanguy Gernot、Emmanuel Giguet和Christophe Rosenberger指导的硕士一年级实习。2025年3月3日至7月4日。[实习报告]

Mohammed Salah AISSAOUI - 4个月。当代作家数字档案调查。文档自动分类。由Tanguy Gernot、Emmanuel Giguet和Christophe Rosenberger指导的硕士一年级实习。2025年3月3日至7月4日。[实习报告]

示意图：

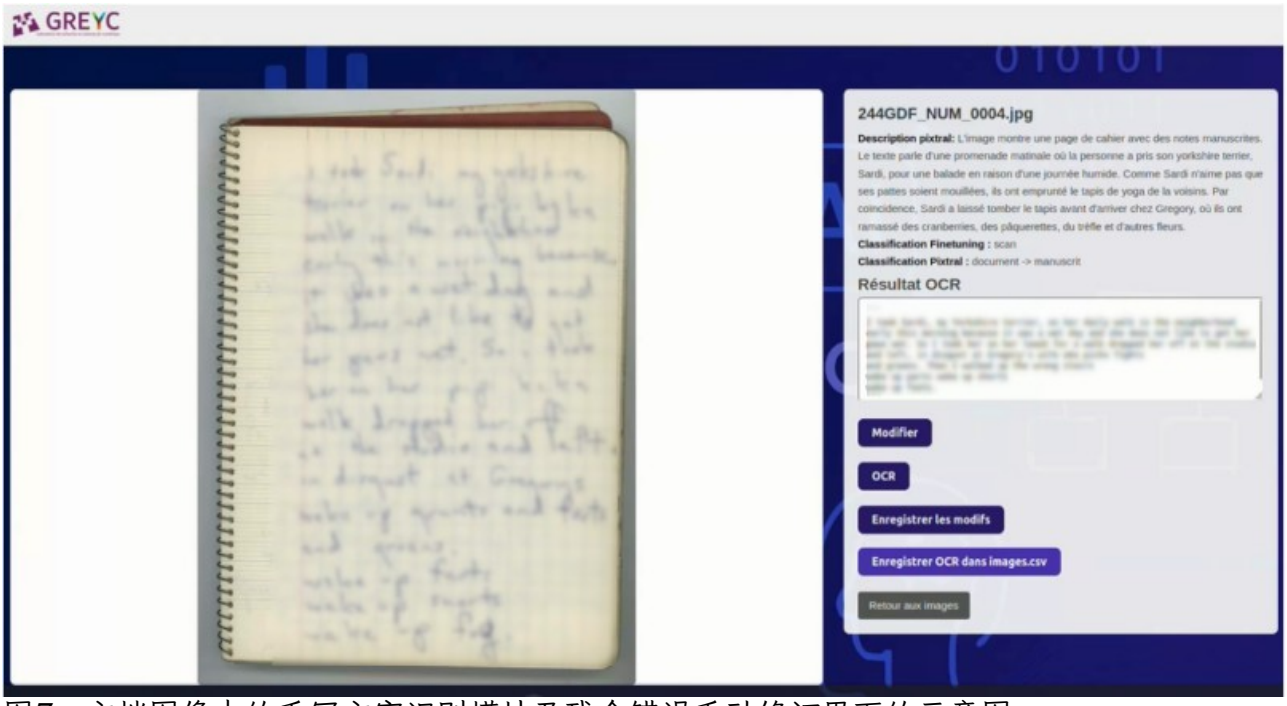
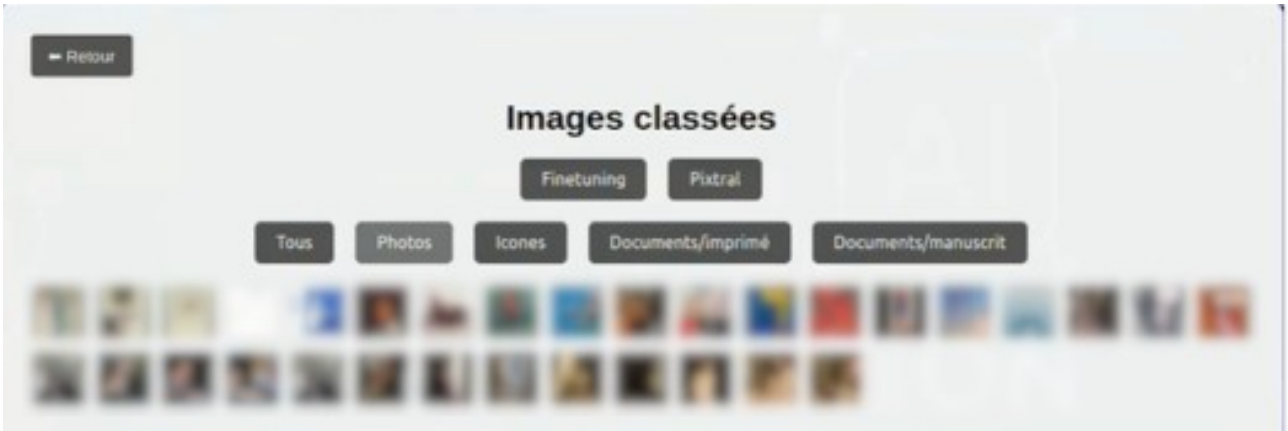


图7：文档图像中的手写文字识别模块及残余错误手动修订界面的示意图



*Figu*图*8*：文档图像自动分类模块的示意图 *s*

7. 仇恨言论检测与档案可传播性

在分析数字档案馆藏时，IMEC 的档案工作者可能会遇到有时较为敏感、甚至可能令人不适的内容。虽然显然有些内容不应被提供或披露，但在档案工作者接触这些内容之前，提前告知其存在似乎同样重要，以便限制因意外遭遇这些内容而可能产生的心理影响。

敏感内容的检测与这些内容是否可以向公众或研究作者档案馆藏的研究人员开放这一问题密切相关。事实上，内容的可传播性这一概念范围更广，涵盖了法规、法律、伦理和合同等多个层面的含义。从本质上说，可传播性并非能够完全通过自动化方式检测的概念。对于处理私人档案的 IMEC 而言，文件是否可传播取决于档案工作者的专业判断，他们最能把握具体的法律背景以及相关作品所涉及的关系。

在此背景下，文本内容自动分析可以通过预先识别侮辱性、歧视性或仇恨言论，为档案工作者的工作提供支持。本项目的目标正是实现这一点：开发一款仇恨言论检测工具，以促进档案管理工作的合乎伦理与安全。项目考虑了 11 类敏感内容：侮辱性言论、仇恨言论、种族主义、宗教、性别歧视、反LGBTQ言论、残障、仇外心理、政治、暴力、威胁。这一分类体系中的部分类别彼此重叠，因此有助于识别无法明确归入某一特定类别的言论。

为实现这一目标，我们开展了多个语料库构建步骤：从档案全宗中提取并准备文本，构建专门用于仇恨言论检测的法语和英语数据集，通过机器翻译扩充语料库，并利用语言模型进行受控生成，以平衡少数类别。我们考察了两种基于Transformer的模型：适用于法语的 CamemBERT，以及稳健的多语言模型 XLM-RoBERTa。这些模型针对 11 个敏感内容类别进行了多标签分类微调。

这项工作建立了一套完整流程，涵盖从数据准备到评估用于档案中仇恨言论检测的 Transformer 模型。然而，由于训练模型的数据不足，结果质量仍受影响，因此尚未能开发出可实际使用的系统。

参考文献：

Hamza OUAZZANI CHAHDI - 4个月。当代作家数字档案取证。当代作家档案的可传播性。由 Christophe Rosenberger 指导的二年级实习。2025年4月15日至8月15日。[实习报告]

示意图：

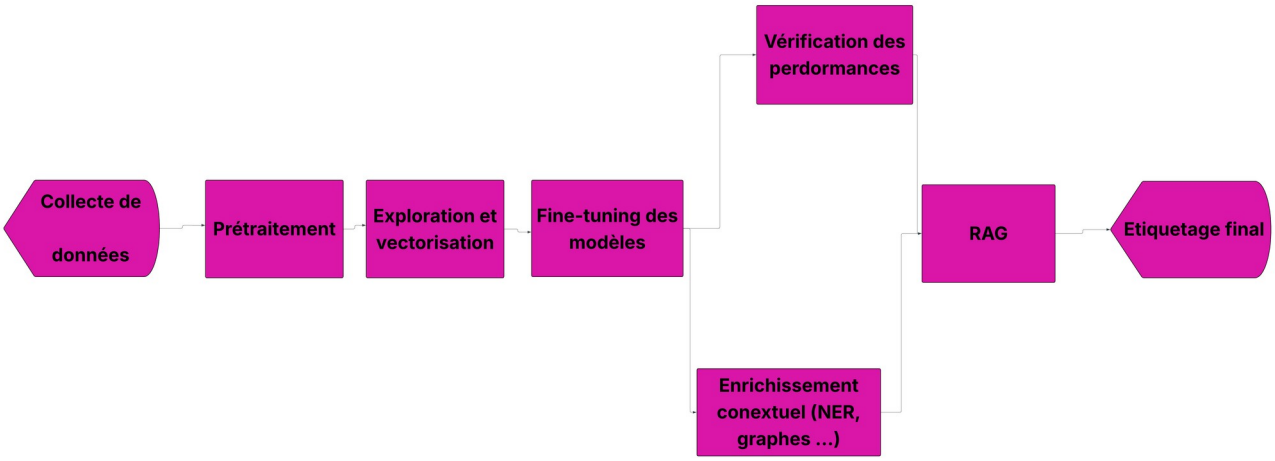


图 *9*：用于仇恨言论检测的处理流程示意图

总结

GREYC 在 IANEC 项目框架下开发的工具均已在 IMEC 的档案数据上经过测试，能够有效辅助数字挖掘档案工作。尤其值得一提的是：

- 一款 Mac 多操作系统仿真工具已多次应用于 IMEC，如今无需处理这些环境复杂的配置，即可恢复旧格式中的文件。

- 一种解决方案可直接访问某些磁盘映像（HFS）的内容，无需借助专用方案，从而简化并自动化对这些映像的批量处理。

- 按文献类型分类的相关工作已在多个档案全宗中成功测试，并且可轻松用于浏览器工具。这些工具可辅助档案管理员开展归档前的初步识别工作，以及识别不可公开的内容元素；

- 应用于“文档图像”类别的分类方法已在 IMEC 的语料库上进行测试，可用于探索在大型语料库中自动提取识别信息，尤其是借助光学字符识别处理手写文档，并自动生成内容描述。

这项工作的后续发展方向很多。首先，可以将项目的交付成果更好地整合到 IMEC 的档案处理流程中，以便在更多样的档案全宗上验证相关工作，同时产出符合档案标准的成果。

基于大型语言模型和人工智能开展的文献分类、内容提取或文档摘要研究已展现出巨大潜力，如今也使人们得以设想以新的方式处理档案。不过，要充分发挥这一潜力，还需要加强 IMEC 的硬件基础设施。此外，也应进一步思考这可能给相关职业带来的方法论变革。

GREYC指导实习

8. 2024年指导实习

Arthur LOMBARDO——2个月。改进 G'DIP 数字取证平台的用户界面。由 Emmanuel Giguet 和 Tanguy Gernot 指导的硕士一年级实习。2024年5月27日至7月19日。

Hamid BEN OMAR - 4个月。将G'DIP取证模块集成到Autopsy中，并为在旧款Mac OS上开展数字取证做好准备。由Emmanuel Giguet和Tanguy Gernot指导的二年级实习。2024年4月15日至2024年8月19日 2024年。 [实习报告]

Thomas VARIN - 3个月。档案图像中的人脸检测与识别。由Emmanuel Giguet和Tanguy Gernot指导的二年级实习。2024年6月3日至8月23日。 [实习报告]

9. 2025年指导实习

Mohammed Salah AISSAOUI - 4个月。当代作家数字档案调查。文档自动分类。由Tanguy Gernot、Emmanuel Giguet和Christophe Rosenberger指导的硕士一年级实习。2025年3月3日至7月4日。 [实习报告]

Matthias DAVID - 4个月。当代作家数字档案调查：为当代出版记忆研究所提供创新数字服务。由 Tanguy Gernot、Emmanuel Giguet 和 Christophe Rosenberger 指导的硕士一年级实习。2025年3月3日至7月4日。[实习报告]

Guillaume HAUTOT - 4个月。当代作家数字档案调查。由 Tanguy Gernot、Emmanuel Giguet 和 Christophe Rosenberger 指导的硕士一年级实习。2025 年 3 月 3 日至 7 月 4 日。[实习报告]

Titouan LE BRET - 4个月。当代作家数字档案调查：数字取证与数字档案处理。由 Tanguy Gernot、Emmanuel Giguet 和 Christophe Rosenberger 指导的硕士一年级实习。2025年3月3日至7月4日。[实习报告]

Hamza OUAZZANI CHAHDI - 4个月。当代作家数字档案调查。当代作家档案的可传播性。在 Christophe Rosenberger 指导下进行的二年级实习。2025年4月15日至8月15日。[实习报告]

Dihia SLIMANA - 4个月。当代作家数字档案调查。文档自动分类。在 Tanguy Gernot、Emmanuel Giguet 和 Christophe Rosenberger 指导下进行的硕士一年级实习。2025年3月3日至7月4日。[实习报告]