



迈向智慧图书馆: 借助 AI 和数据在未来规划中的应用

Smart Library: Planning the Future with AI and Data

顾锦明 (香港大学图书馆 数码发展及支持服务助理馆长)

Dr. Ku Kam Ming (Associate Librarian, Technology Support Services, HKUL)

马颖怡 (香港大学图书馆 数据服务及信息研究主任)

Ms. Mahboobani Vanessa Ramesh (Data Services & Information Research Librarian, HKUL)

智慧图书馆 Smart Library

“智慧图书馆”一词是由Aittola、Ryhanen和Ojala [1] 于2003年首次提出的。

The term "smart library" was first introduced by Aittola, Ryhanen, and Ojala in 2003.

智慧图书馆的目标是提升图书馆内所有元素和活动的智能化，从图书馆的日常管理到用户服务，甚至包括后台的工作。

A smart library aims to enhance the intelligence of all elements and activities within the library, from housekeeping operations to user services, including the behind-the-scenes activities.

实体应用案例

- **数字图书馆和资料库**：提供电子书、在线期刊、数据库和特殊收藏的访问平台。
- **移动应用**：允许用户通过智能手机搜索目录、借阅资料以及访问数字资源的应用程序。自助借还机和账户管理设备，减少等待时间，提升便利性。
- **RFID技术和自助服务终端**：用于追踪库存和便利自助借阅流程的射频识别技术。
- **物联网 (IoT) 设备**：用于监测图书馆环境、管理能源使用和提升用户体验的物联网设备。

实体应用案例

- **智能家具**：配备技术的家具，如充电站、互动显示屏和灵活使用的模块化设计。
- **虚拟现实（VR）和增强现实（AR）**：提供沉浸式学习、探索和互动展览体验的技术。
- **云计算**：通过互联网访问的存储和服务，允许可扩展资源和远程数据访问。

怎样的图书馆建设才是有智慧？

不同学者有不同的看法。

Different scholars have different perspectives.

智慧的一种定义：

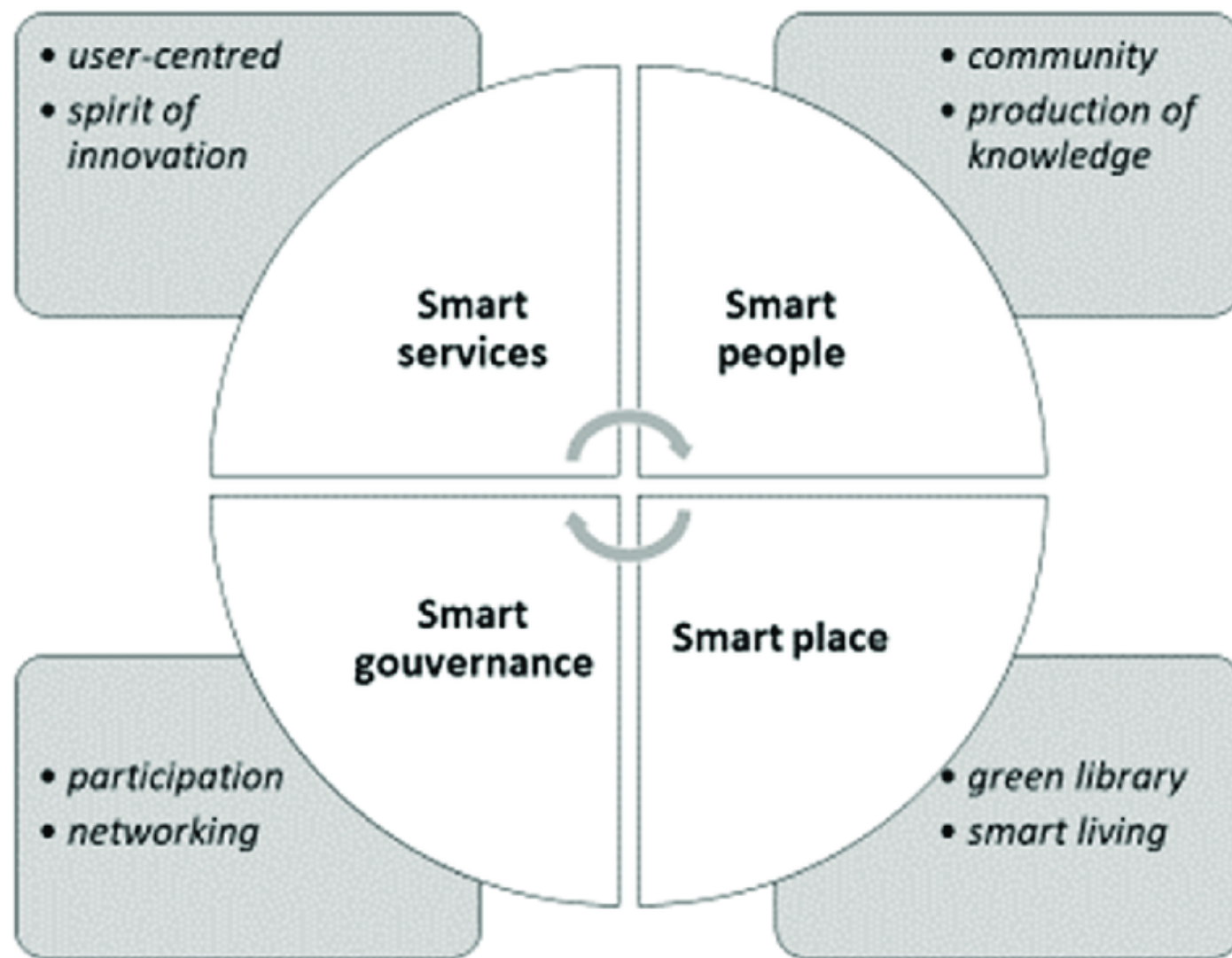
智慧指的是适应、学习和有效应对变化环境和用户需求的能力。它通常涉及技术的整合，以增强功能和效率。

怎样的图书馆建设才是有智慧？

Sotonye Orji 和 Isaac Echezonam Anyira [2] 提出

- S = Services 服务
- M = Methods 方法
- A = Automation 自动化
- R = Resources 资源
- T = Technology 科技

Schöpfel, J. (2018)
的理解有四個項目。



Zimmerman 和 Chang (2018) [4] 对智慧图书馆的看法如下：智慧图书馆是书籍及相关数据与数字资源和空间资源的整合，基于图书馆全面信息化的实现。通过使用大数据智能分析平台，图书馆的管理和服务变得智能化和个性化，以满足读者需求，提升他们的体验。为了实现这一目标，智慧图书馆的建设需要真正实现全面的数据交换，并建立一个全面的数据共享中心。

A smart library integrates books and **related data along with digital and spatial resources**, founded on the comprehensive digitization of library services. By utilizing a **big data intelligent analysis platform**, library management and services become intelligent and personalized, catering to the needs of readers to enhance their experience. To achieve this, the construction of a smart library necessitates the realization of true full data exchange and the **establishment of a comprehensive data-sharing center.**

人工智能在智慧图书馆中的角色

在智慧图书馆的背景下，人工智能通过数据分析、学习和决策过程模仿人类智慧，预测趋势，提供个性化服务(例如 个性化推荐、用于协助的聊天机器人)并优化运营，增强了图书馆服务社群的能力。

人工智能并可以通过自动化日常任务和提供资源配置的洞察，提升了运营效率和资源管理。

随着技术的发展，人工智能与图书馆智慧之间的关系将更加紧密。机器学习和自然语言处理等新兴技术将进一步增强用户体验和运营能力。

数据对人工智能至关重要

- **学习基础**：人工智能系统依赖大型数据集来学习模式和进行预测。数据的质量和数量直接影响人工智能模型的准确性和有效性。
- **训练与验证**：机器学习算法需要多样且具代表性的数据来有效训练。不足或偏见的数据可能导致性能不佳和结果失真。
- **持续改进**：人工智能模型需要持续的数据输入，以便随时间适应和改进。定期更新增强了它们对新趋势和用户行为的响应能力。
- **决策支持**：数据驱动的洞察使人工智能能够做出明智的决策和推荐，确保在各种应用中的相关性和准确性。

数据是成功的关键

Data is the key to success

- 没有数据 ≠ 没有人工智能

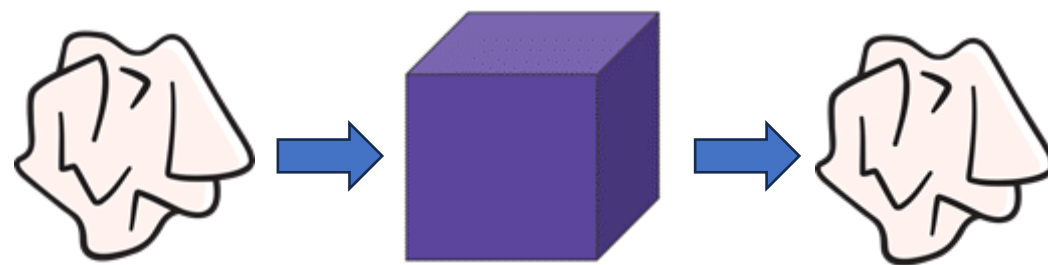
No data ≠ No AI

- 人工智能模型基于数据进行学习和训练
AI model learns from and train based on data

- 垃圾进，垃圾出

Garbage In, Garbage Out

- 更多洁净的数据 = 更好的结果
More **clean** data = better results



数据对智慧图书馆的重要性

- **理解用户需求：** 收集用户互动和偏好的数据帮助图书馆定制服务和资源，以更好地满足社区需求。
- **资源分配：** 通过分析数据，图书馆可以优化其藏书和服务，确保资源分配到最需要的地方。
- **项目开发：** 数据驱动的洞察使图书馆能够根据用户兴趣和参与趋势设计项目和活动，从而增强社区参与感。
- **绩效评估：** 图书馆可以利用数据评估服务和举措的有效性，促进持续改进和问责制。
- **个性化体验：** 通过利用用户数据，智慧图书馆可以提供个性化推荐和定向沟通，丰富用户体验。

工作案例

如何构建数据分析

How to do data analysis

如何构建数据分析

How to build data analysis



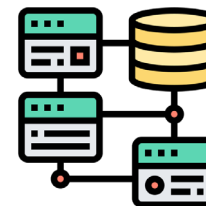
定义目标
Define objectives



收集数据
Data collection



数据整理
Data preparation



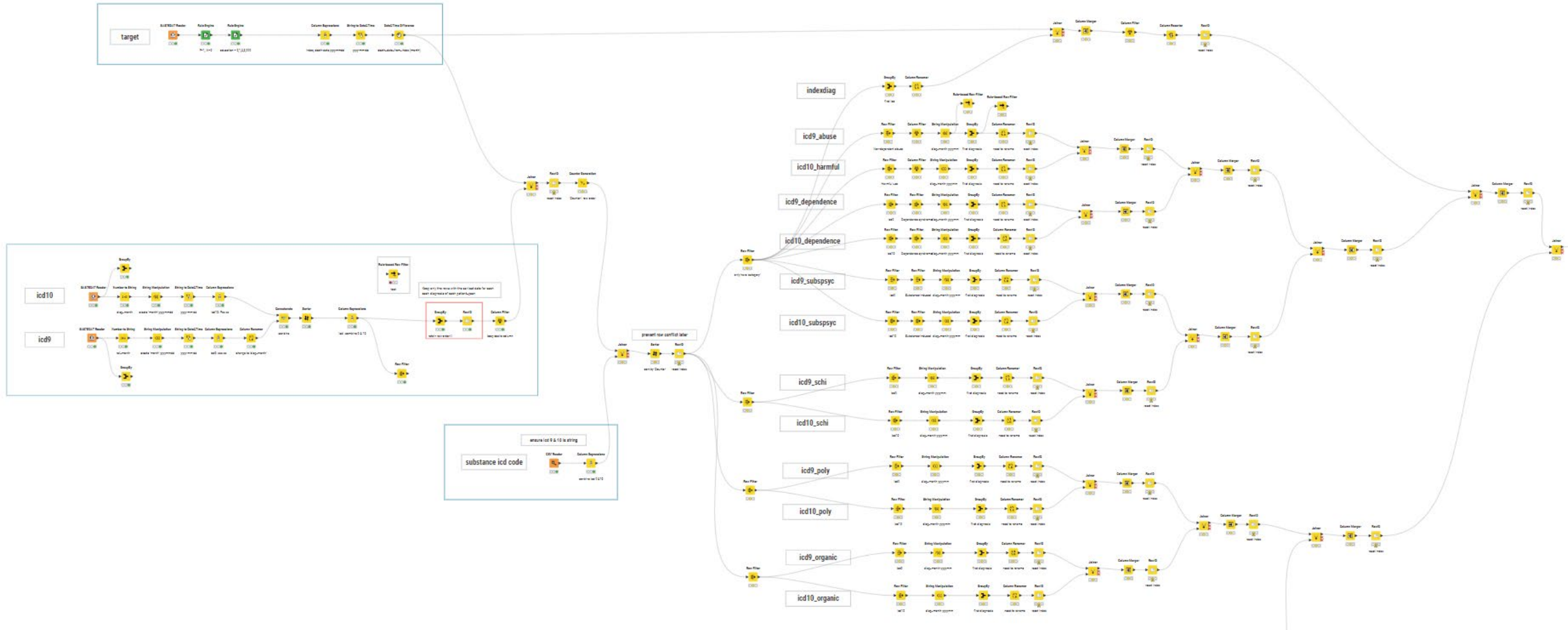
数据分析
Data Analysis



可视化结果
Visualize Findings

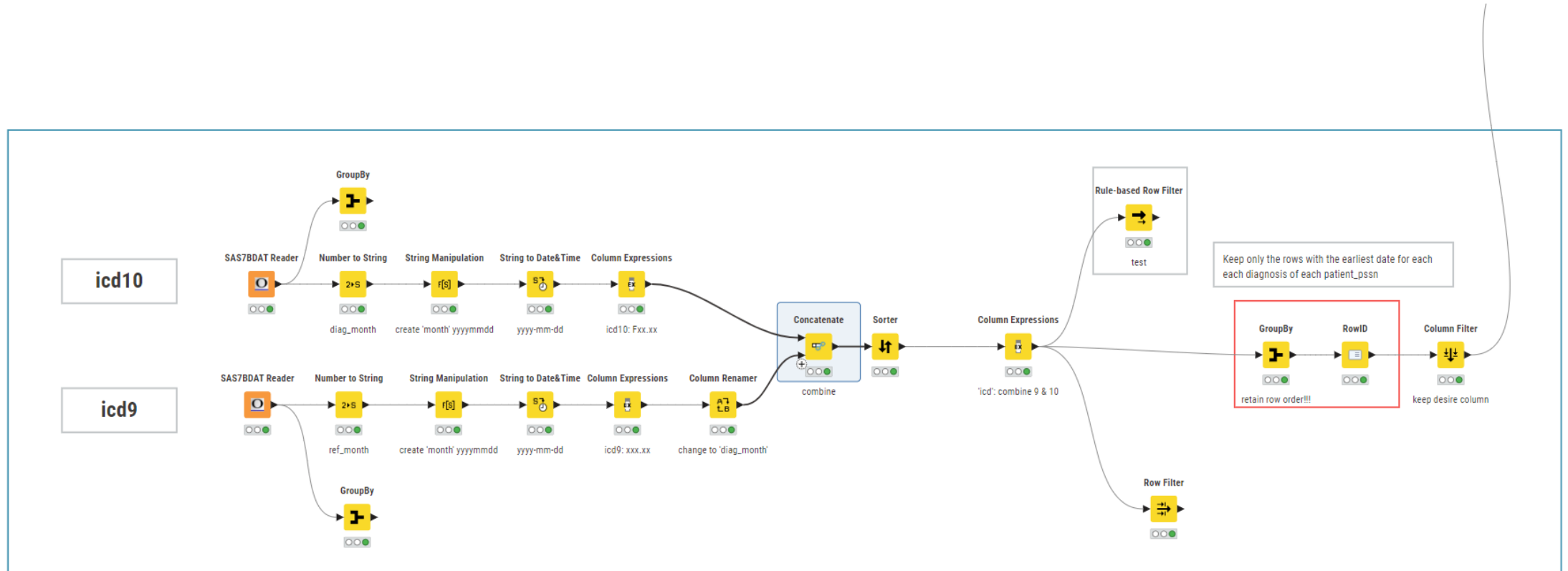
KNIME 工作流程

KNIME workflow



KNIME 工作流程

KNIME workflow



案例研究 1

Case Studies 1

电子资源使用情况分析报告

Electronic resources usage statistics

电子资源使用情况分析报告

Electronic resources usage statistics

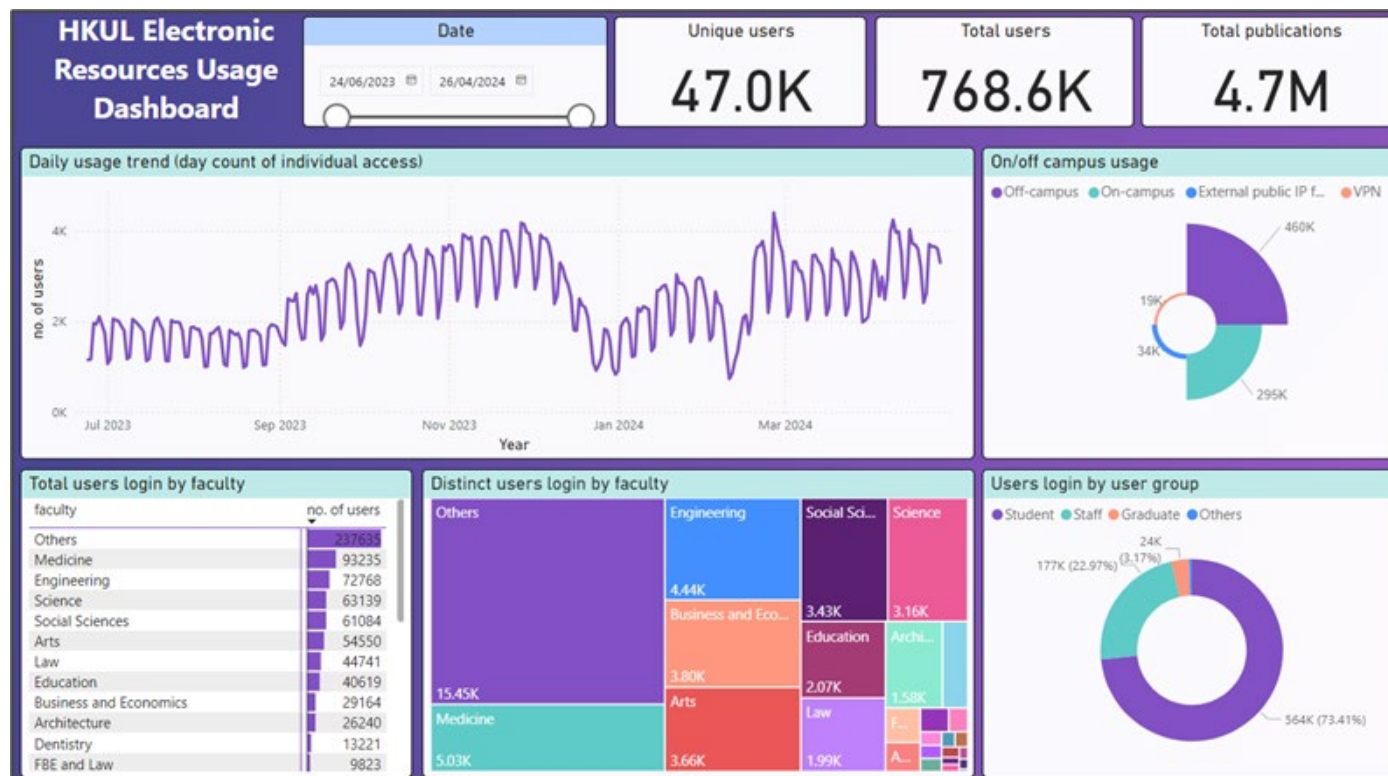
从 EZproxy 日志取得电子资源使用数据及加以分析

Get e-resources usage from EZproxy log and analysis the data

识别及深入了解不同使用群组的使用量
Identify user unique behaviours and usage by different user group

适切了解不同群组不断变化的需求

Better understand the changing demands of user group



电子资源使用情况分析报告

Electronic resources usage statistics

发现隐藏中使用模式和偏好

Uncover hidden usage patterns and preferences

例子 Example

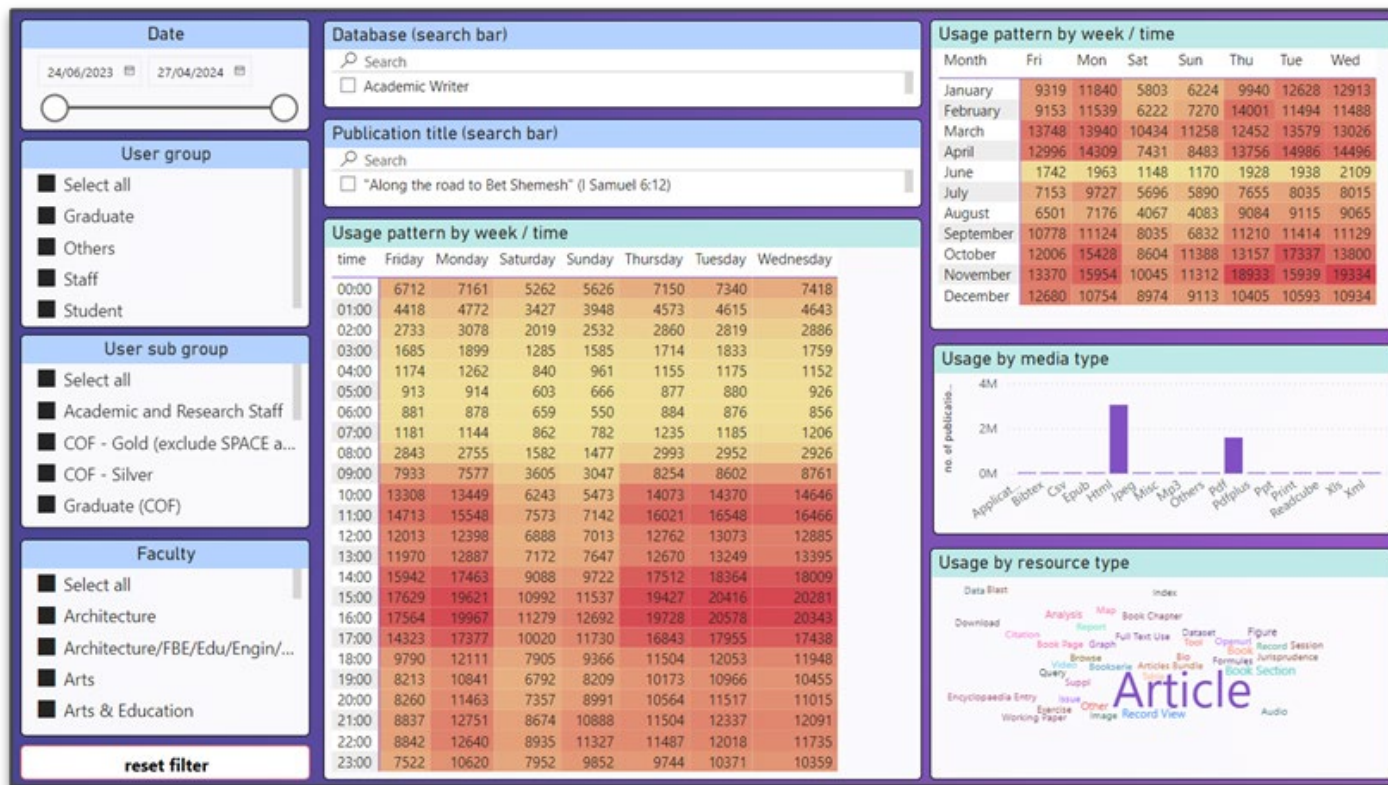
高使用时段：每年的3-4, 10-11月及每日的10:00 - 18:00；

最多人使用资源类型：文章

High usage periods :

Mar.-Apr., Oct.-Nov. and 10:00-18:00
every day;

Most used resource type : Article

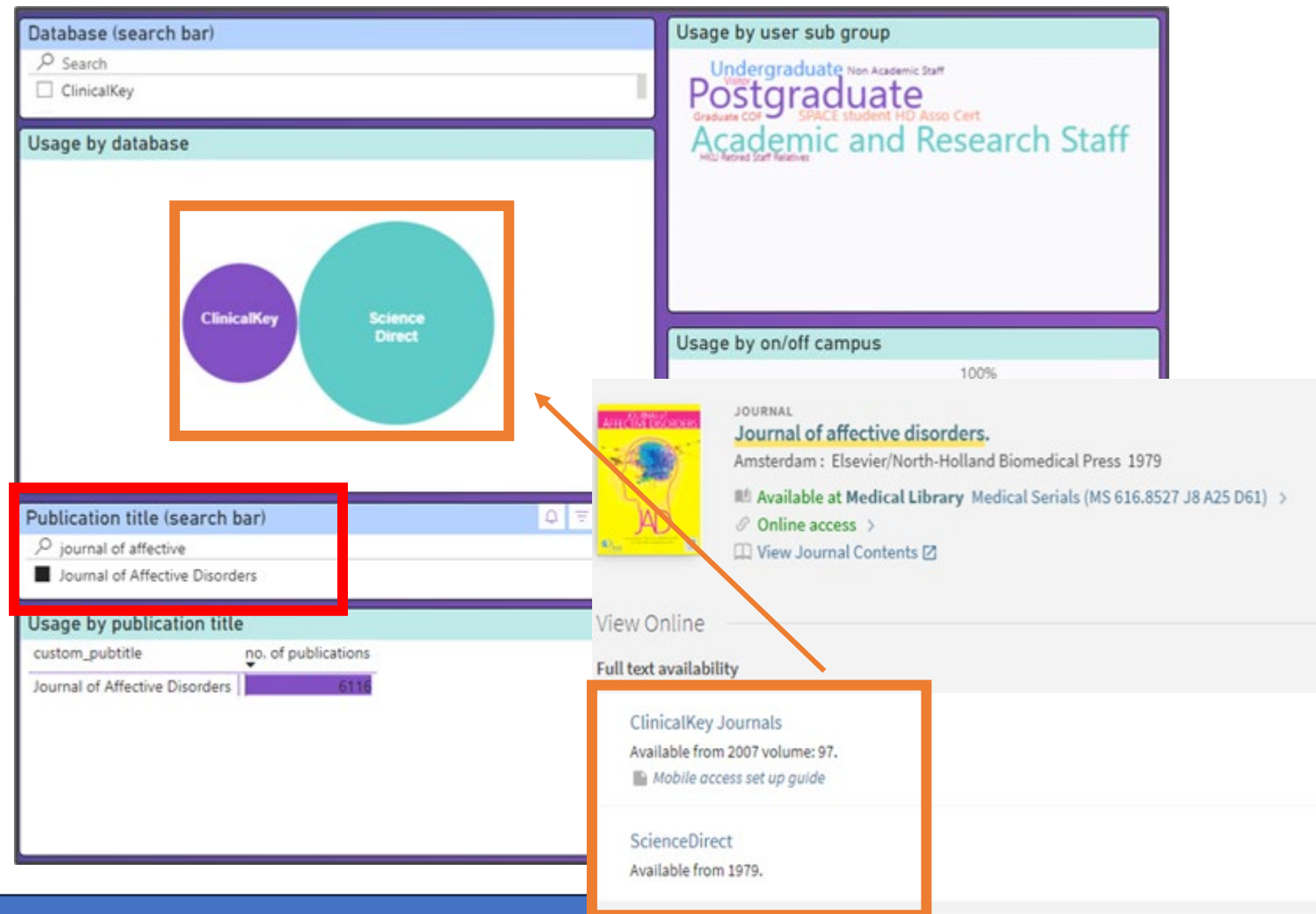


电子资源使用情况分析报告

Electronic resources usage statistics

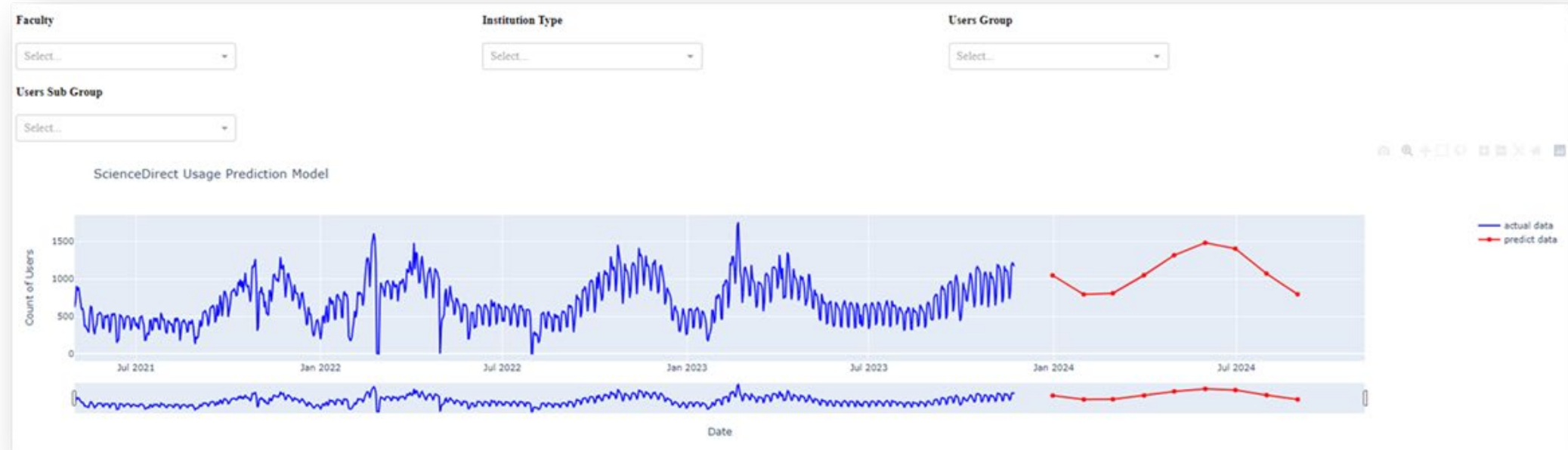
识别相同出版物在多个数据库的使用量，帮助管理层在图书馆服务和资源订阅及分配方面做出决策

Identify usage of the same publication across multiple databases to help management make decisions on library services and resource subscriptions and allocations



电子资源使用情况分析报告

Electronic resources usage statistics



使用趋势预测
Usage forecasting

作出基于证据的决策
Evidence-based decision

制定数据驱动策略
Develop data-driven strategies

电子资源使用情况分析报告

Electronic resources usage statistics

使用GenAI 技术直接与数据 "谈话"
interact with data
conversationally using GenAI

即时的数据摘要帮助管理层快速做出明智的决策

Immediate data summaries help management to make informed decisions quickly

Which platform should I continue subscribe?

To decide which platform to continue subscribing to, consider the following factors:

- 1. Usage/Access Frequency:**
 - **Science Direct** shows the highest usage (358 occurrences), suggesting that it's the most accessed platform for "The Lancet."
 - **The Lancet** platform is used 123 times, while **ClinicalKey** has only 19 occurrences.
- 2. Content Needs:**
 - If Science Direct provides sufficient access to the content of "The Lancet" that your users need, it could be the primary platform.
 - Consider whether The Lancet platform itself provides exclusive content or additional features beyond what's available on Science Direct.
- 3. Cost Efficiency:**
 - Compare subscription costs for each platform. If Science Direct covers your needs comprehensively, and at a lower cost, it might be more economical to focus on that platform.
- 4. User Preferences:**
 - Analyze any feedback or preferences from the users or departments accessing these platforms. A platform with a lower access rate might still be important to a specific group.

Recommendation:

Based on the usage data, **Science Direct** should be prioritized for continued subscription as it has the highest access rate. You could consider discontinuing **ClinicalKey** due to its low usage unless it provides essential content not available on the other platforms.

提出问题
应该继续订阅哪个平台？

数据摘要及分析过程

1. 使用率
2. 内容需求
3. 成本效益
4. 用户偏好

结论及决策

Science Direct使用率最高应继续订阅，可考虑停止订阅使用率很低的ClinicalKey

案例研究 2

Case Studies 2

研究主题趋势分析

Research topic trend analysis

研究主题趋势分析

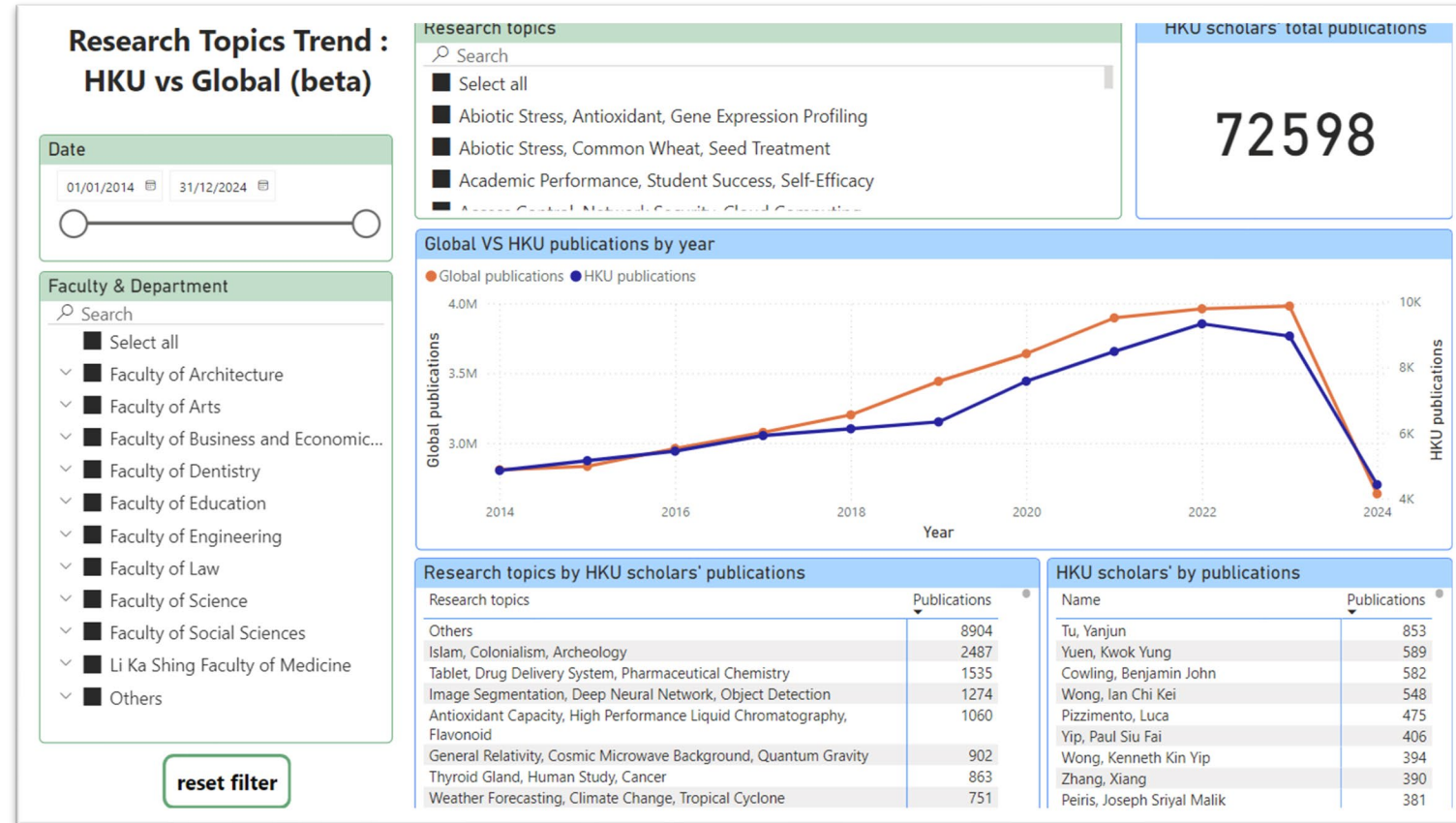
Research topic trend analysis

分析本校与世界的研究主题趋势

compare institution's performance and focus research areas with global institution

识别及深入了解不同学院, 部门及学者的研究主题趋势

Identify faculty, department and scholars invloved in top emerging research topic

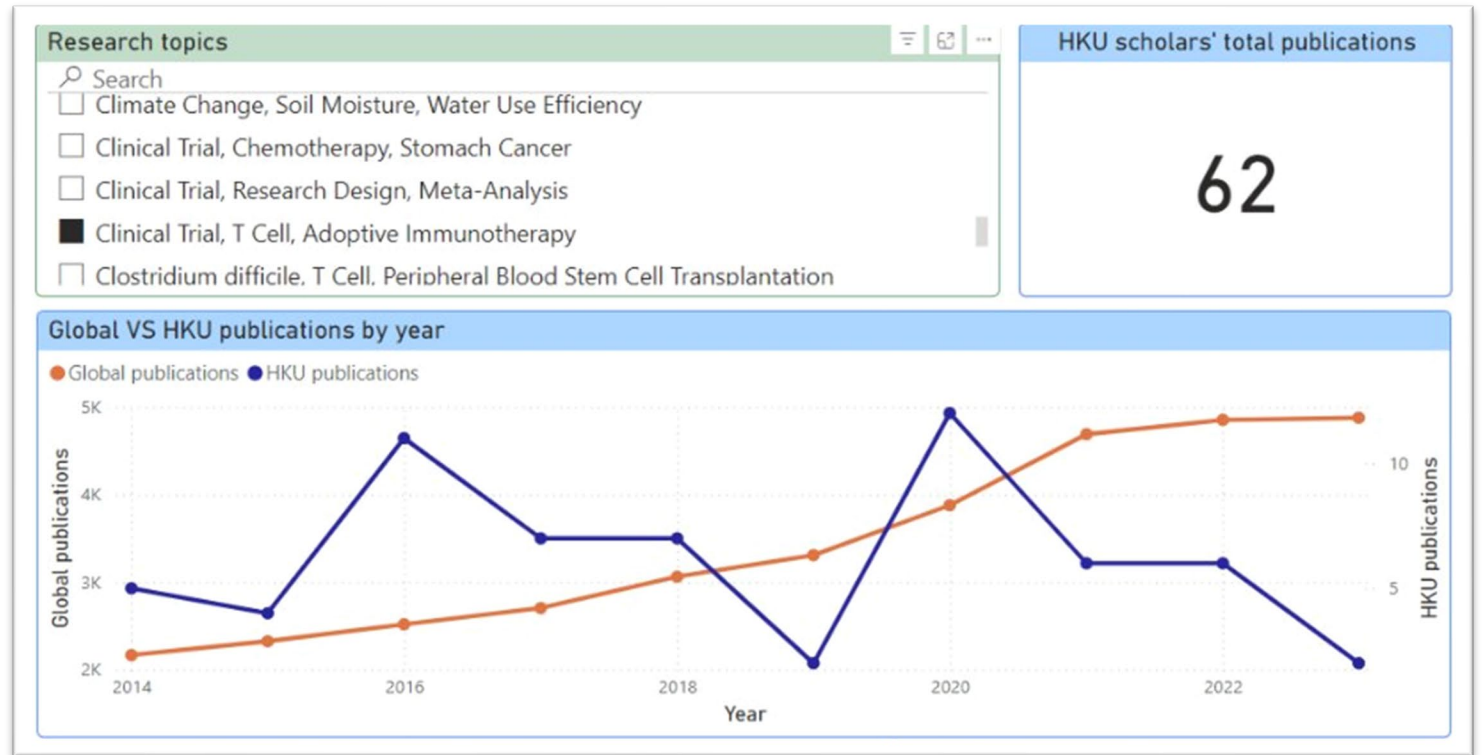


研究主题趋势分析

Research topic trend analysis

识别本校在全球趋势中落后的领域，帮助大学管理层在未来研究方向、资金和人才需求上做出以数据为中心的决策

Identify gaps where institution lag behind global trend which help management in making data-centric decisions on future research directions, funding, and talent requirement



研究主题趋势分析

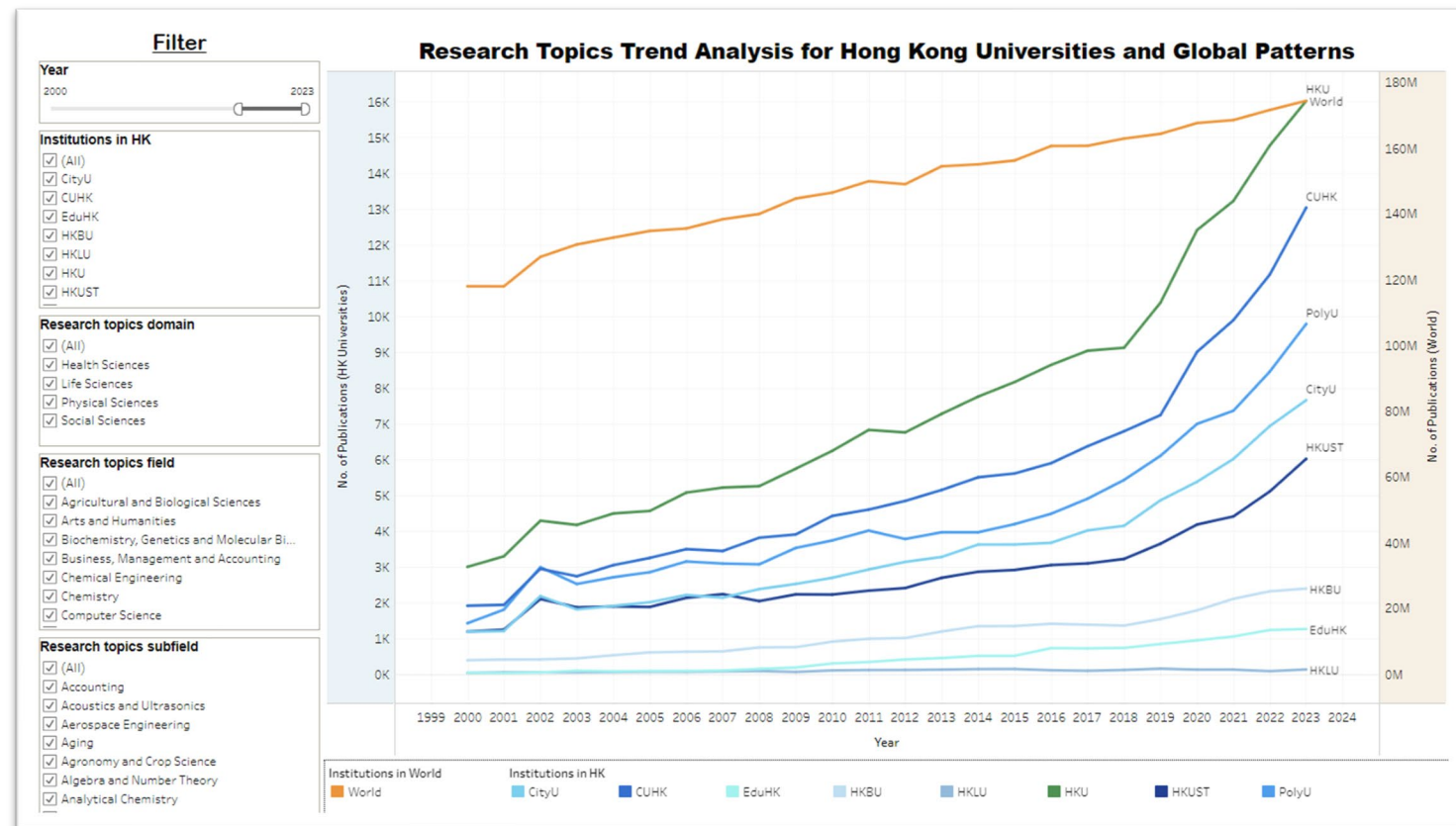
Research topic trend analysis

将本校的研究重点领域与本地同行及全球机构进行比较

compare institution focus
research areas with local peer
and global institution

与同行机构进行基准比较，
以确定竞争优势和差距

Benchmark against peer
institution to identify
competitive advantages and
gaps



研究主题趋势分析

Research topic trend analysis

使用GenAI 技术直接帮助管理层提升了决策的准确性，还大幅节省了管理层的时间，使他们能够专注于更具战略性的任务

Utilize GenAI technology to help management improve the accuracy of decision-making as well as save their time which allow them to focus on more strategic plan and tasks

Which subfield that World is increasing over years but HKU is decreasing over year?

The subfields where the World is showing an increasing trend while HKU is showing a decreasing trend are those where:

- The publication count for the **World** increases year over year.
- The publication count for **HKU** decreases year over year.

I have already analyzed and identified several subfields with opposite trends. To clarify, here are the **top 5 subfields** where this condition holds:

1. **Complementary and Alternative Medicine** (+2586)
2. **Otorhinolaryngology** (+2145)
3. **Computer Graphics and Computer-Aided Design** (+1933)
4. **Applied Mathematics** (+1284)
5. **Biochemistry** (+860)

提出问题

哪个领域世界在逐年增加而香港大学却逐年减少？

直接结论

- 1 补充和替代医学
2. 耳鼻咽喉科
3. 计算机图形学和计算机辅助设计
- 4.应用数学
- 5.生物化学

使用 AI 和数据帮助未来规划



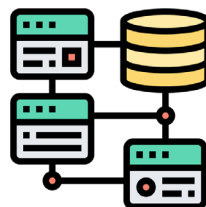
定义目标
Define
objectives



收集数据
Data collection



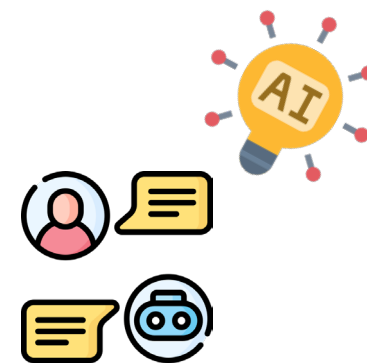
数据整理
Data
preparation



数据分析
Data
Analysis



可视化结果
Visualize
Findings



得出结论及决策建议
Draw conclusions
and make decisions

总结

数据是人工智能和智慧图书馆的支柱。对于人工智能而言，数据对于模型训练和做出明智决策至关重要。对于智慧图书馆，数据是理解用户需求、优化资源和提升整体服务质量的关键。有效利用数据使得人工智能技术和图书馆系统能够不断发展，以满足用户和社区不断变化的需求。

Data is the backbone of both AI and smart libraries. For AI, it is crucial for training models and making informed decisions. For smart libraries, data is vital for understanding user needs, optimizing resources, and improving overall service delivery. The effective use of data enables both AI technologies and library systems to evolve and respond to the changing demands of users and communities.

文献参考References

- [1] Aittola, M., Ryhänen, T., Ojala, T. (2003). SmartLibrary – Location-Aware Mobile Library Service. In: Chittaro, L. (eds) Human-Computer Interaction with Mobile Devices and Services. Mobile HCI 2003. Lecture Notes in Computer Science, vol 2795. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-45233-1_38
- [2] Orji, S., & Echezonamanyira, I. (2021). What is “Smart” About Smart Libraries? International Journal of Research in Library Science, 7(4), 265. <https://doi.org/10.26761/ijrls.7.4.2021.1482>
- [3] Schöpfel, J. (2018). Smart libraries. Infrastructures, 3(4), 43. <https://doi.org/10.3390/infrastructures3040043>
- [4] Zimmerman, T., & Chang, H.-C. (2018). Getting Smarter: Definition, Scope, and Implications of Smart Libraries. *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*, 403–404. <https://doi.org/10.1145/3197026.3203906>

谢谢!

Thank you!

