

活用数字数据构建研究平台： 研究人员与图书馆员协作下的 “数字源氏物语”尝试^{*}

中村美里 木越みち 小川夏代子 王子睿/译

摘 要 东京大学附属图书馆在对馆藏《源氏物语》抄本进行数字化之时,组织了由不同领域研究者和图书馆员组成的学习会,围绕《源氏物语》的数字数据应用展开研讨。作为成果之一,学习会构建并发布了一个旨在为《源氏物语》文本研究提供支持的平台——“数字源氏物语”。文章将阐述学习会成立之缘起、平台构建及功能拓展之历程,并分别探讨学习会本身及图书馆员与研究人员协作之意义。

关键词 数字档案; 数字人文; 人文信息学; IIR; 教职合作; 源氏物语

分类号 G35; TP39

作者简介 中村美里(Nakamura Misa),东京大学附属图书馆信息服务部资料整備组主任,E-mail:nakamura.misa@mail.u-tokyo.ac.jp; 木越みち(Kigoshi Michi),东京大学附属图书馆信息管理部信息管理组主任(负责资料合同); 小川夏代子(Ogawa Kayoko),东京大学附属图书馆总务部企划外联组主任; 王子睿(译者),北京语言大学博士研究生,E-mail:oushieiwzr@gmail.com。

1 简介

大学图书馆在进行馆藏文献数字化公开之时,除了构建和维护数字档案馆系统、创建图像及其创建元数据外,如何灵活应用数字数据也被视为其中一个重要的课题。东京大学(以下简称“本校”)附属图书馆^①以本校综合图书馆所藏《源氏物语》抄本(以下简称“东大本”)数据为核心,持续举办研究人员与图书馆员共同参与的学习会,以探讨数字数据的灵活运用。作为学习会的成果之一,创建与公开了旨在为《源氏物语》文本研

^{*} 原文“研究プラットフォーム構築のためのデジタルデータ活用——教職協働による「デジタル源氏物語」公開の試み”发表于2024年日本期刊《大学图书馆研究》126卷,2024年11月12日获得论文作者、项目团队以及期刊编委授权本刊中文翻译出版许可。——编者

^① 东大校内三十余所图书馆及阅览室统称为“附属图书馆”,其中以位于本乡校区、规模最大的“综合图书馆”为首。

究提供支持的平台——“数字源氏物语”,并在公开后持续进行了功能拓展。本文首先概述了《源氏物语》研究及其相关研究文献,继而详述东大本及其数字化的过程。其后,将具体阐述与研究者共同举办学习会的始末,“数字源氏物语”创设之缘起,平台发布前的准备工作以及发布后的功能拓展等。最后,本文亦将探讨学习会本身以及图书馆员与研究人员协力合作之价值。

2 《源氏物语》研究及相关研究资料

《源氏物语》为紫式部创作于日本平安时代中期的长篇小说。该作品以主人公光源氏及其周围人物的际遇为线索,描绘了彼时平安贵族的生活、社会以及思想等。据传,《源氏物语》抄本主要可分为三种:青表纸本(藤原定家校订本系统)、河内本(河内守源光行、源亲行父子校订本系统),此外尚有其他抄本,并存有诸多版本。平安时代末期迄今的学者们撰写了大量《源氏物语》注疏、相关著作及研究论文。本节将重点介绍与“数字源氏物语”相关的两部文献:《校异源氏物语》(全5卷,芳贺博士纪念会编,中央公论社出版,1942)及《新编日本古典文学全集》(全88卷,小学馆出版,1994—2002)。

《校异源氏物语》(以下简称《校异》)为日本国文学者池田龟鉴(1896—1956)所著之《源氏物语》研究著作。该书详细记录了池田龟鉴对青表纸本、河内本等多种《源氏物语》抄本的语句、文字异同进行文本勘校的结果。此外,《校异》刊行后,池田龟鉴又进行了增补修订,并将其收录于《源氏物语大成》(全8卷,池田龟鉴编,中央公论社出版,1953—1956)之中。《源氏物语大成》将修订后的《校异》以“校异篇”(卷1—3)的形式收录其中,并增补了索引篇(卷4—6)、研究·资料篇(卷7)及图录篇(卷8)。

《新编日本古典文学全集》(以下简称“《新编》”)收录了日本古代至江户时期的文学作品,并附有原文及现代日语译文。《源氏物语》收录于《新编》第20卷至第25卷,其底本主要为传定家笔本、传明融笔临摹本、飞鸟井雅康笔本(日本古代学协会藏,通称“大岛本”)等。

3 东大本及其数字化

3.1 什么是东大本^①

本文所称“东大本”,系本校综合图书馆所藏之珍贵典籍《源氏物语》抄本54册(索书号A00:6587)。在

^① 本节及下节内容参考了以下在线文献:田村隆《东京大学综合图书馆藏〈源氏物语〉》, <https://www.lib.u-tokyo.ac.jp/ja/library/general/genji-handbook>, 访问日期:2023年10月24日。

《源氏物语大成》第7卷、《源氏物语别本集成》第15卷(源氏物语别本集成刊行会编,おうふう出版,2002)及续第7卷(2010)等文献皆对其有所论述,亦被称为“东京大学本”或“东大本”。其撰写年代据推测为室町时代中期,亦有观点认为其成书于室町时代后期或江户时代初期。总观而言,该抄本兼具青表纸本之特征,部分卷册疑似抄录自河内本或其他版本。

东大本为本校综合图书馆所藏青洲文库的一部分。青洲文库是山梨县西八代郡市川大门町(今日本市川三乡町)渡边家三代人所聚之书籍,由寿(1803—1875)、信(1840—1911)及泽次郎(1870—1941)三代积累而成,文库名取自第二代主人信的号“青洲”。1923年(大正十二年)的关东大地震火灾致使东京帝国大学附属图书馆(今综合图书馆前身)文献散佚殆尽,为重振图书馆,馆于1924年(大正十三年)购入青洲文库。青洲文库藏书约25000册,囊括三代渡边家所藏之各类文献;寿所藏之和歌、国学相关资料,信所藏之汉籍、历史书籍以及泽次郎所藏之浮世草子、洒落本等。

青洲文库所藏《源氏物语》除东大本外,尚存传嵯峨本之庆长古活字版(索书号A00:5823),此版本亦被认定为珍贵典籍。

3.2 东大本的数字化

3.2.1 东京大学数字档案馆建设项目

为通过本校所藏学术资源的数字档案馆化以促进资源灵活运用,从而强化支撑学术多元化的基础,东京大学于2016年9月设立“东京大学学术资产等档案馆委员会”。该委员会自2017年度起实施“东京大学数字档案馆建设项目”(以下简称“建设项目”,见图1),旨在将校内多元学术资产等数字化并面向海内外公开,以促进本校资源的广泛利用。此外,为推动建设项目的实施,本校于附属图书馆内设立学术资产档案馆化推进室,主要由图书馆员兼任室员,致力于为校内学术资产的数字化及公开提供支持与后援工作。

建设项目每年均在校内公开征集学术资产等数字化项目,前述档案馆委员会负责对申请内容进行评审与遴选,并对获批项目分配预算支持。原则上,获得预算支持的学部等单位自行负责图像等数字化数据的管理与维护,若相关单位在系统构建与维护方面存在困难,则由建设项目负责管理与维护的图像公开共享服务器进行数字化数据的公开与存储。此外,东京大学设立有致力于数字化学术资产公开的平台“东京大学学术资产及相关档案馆链接集”^①和“东京大学学术资产及相关档案馆门户网站”^②。其中,“东京大学学术资源及相

① 东京大学学术资产等档案馆链接集:<https://da.dl.itc.u-tokyo.ac.jp/dalink/>,访问日期:2023年10月24日。2024年5月末系统更新后,网址已变更为<https://da.dl.itc.u-tokyo.ac.jp/portal/database/list>。

② 东京大学学术资产等档案馆门户网站:<https://da.dl.itc.u-tokyo.ac.jp/portal/>,访问日期:2023年10月24日。2024年5月末系统更新后,网址已变更为<https://da.dl.itc.u-tokyo.ac.jp/>。

档案馆链接集”以列表形式介绍了各学部等的公开数字档案馆网站,同时支持以网站名称为关键词进行检索或根据资源公开单位(如学部)进行筛选。“东京大学学术资产及相关档案馆门户网站”不仅支持对单个资料进行跨库检索,亦提供了通往各数据的便捷链接。此外,该门户网站的检索对象不仅涵盖本项目预算支持所生成的图像及其元数据,还包括各学部等单位独立开展的项目成果,以及由国文学研究资料馆推进的“日本历史典籍国际合作研究网络构建计划”(歴史的典籍 NW 構築計画)产生的数字化学术资产等。关于建设项目,敬请参阅图 1。

3.2.2 东大本数字化历程

东大本最初并未被列为珍贵典籍,而仅作为普通和装本陈列于开放式书库中供师生自由查阅。随后,如 3.1 节所述,经鉴定确认其珍贵价值,最终于 2018 年根据综合图书馆的相关规定被指定为珍贵典籍。当时有观点认为通过本项目对东大本进行数字化公开最为适宜,故同年对东大本进行了全页数字化处理,并于 2019 年 6 月以“东京大学综合图书馆藏《源氏物语》”^①的形式公开发布。公开时,特邀本校田村隆副教授(东京大学研究生院综合文化研究科)撰写了文献解说,并予以一同刊载。

凡注明图像出自本校综合图书馆藏品者,无论商业或非商业用途,皆可自由使用,无需另行申请。

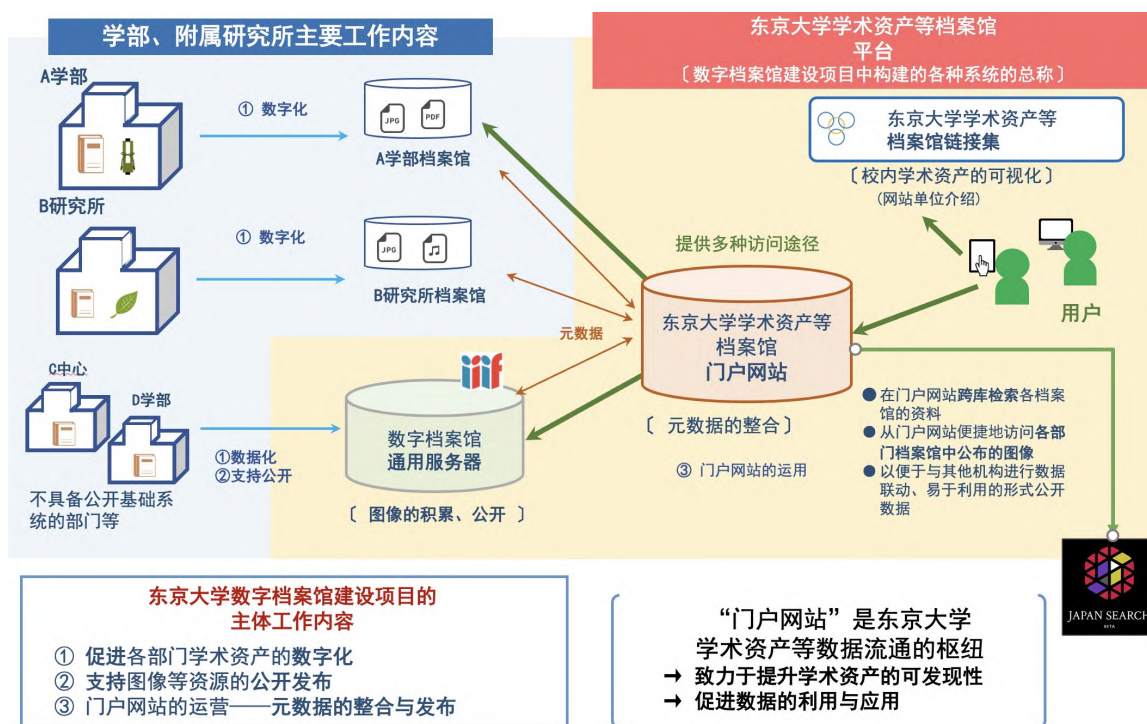


图 1 东京大学数字档案馆建设项目系统概要图

① 东京大学综合图书馆所藏《源氏物语》：<https://iif.dlittc.u-tokyo.ac.jp/repo/s/genji>, 访问日期:2023 年 10 月 24 日。

4 “里源氏学习会”的启动和“数字源氏物语”的发布

4.1 学习会启动缘起

在公开东大本图像的过程中,田村副教授撰写的各种解说和建议使项目得以顺利推进。他深入浅出地向工作人员讲解了《源氏物语》抄本系统及东大本之定位与特色等,并以为之为基础整理成了“资料解说”。此外,根据田村副教授的提议,团队还制作了网页版资料解说,利用 IIIF 的注释功能,并将阅读器嵌入到对应的页面中,方便读者便捷地查阅解说中所提及的内容。

在此次交流的契机下,图书馆馆员提出了以下建议:第一,图像的公开不应为终点,更应探究其在《源氏物语》研究乃至日本国文学领域中的数字化应用潜力,并听取研究者对于未来《源氏物语》数字化工作的期许;第二,希望能与《源氏物语》的研究人员共同探讨数字档案馆应具备的功能。田村副教授欣然接受了上述建议,此外,信息基盘中心助理教授兼档案化推进室助理教授中村觉(2020年7月起兼任史料编纂所助理教授),以及一般财团法人人文信息学研究所永崎研宣主席研究员等学者也对此表达了浓厚的兴趣,并与数位综合图书馆职员共同组建了“里源氏学习会”(以下简称“学习会”)。“里源氏”这一名称,缘于此学习会建立初期为工作时间之外的非正式活动,意在区别于东大本图像公开网站的官方内容(即“表”)——学习会旨在创造一种与之不同、具有实验性和探索性质的东西(即“里”),于是在命名之中赋予了一份幽默。

2019年7月的首次学习会上,田村副教授首先介绍了《源氏物语》研究中的必备图书及重要参考资料,并详述了《校异》的编纂过程、内容细节及其五卷本页码连贯、逐页标注行号的特点,强调这些编号可作为检索关键。他还提到,在教学中,内容详尽、兼具原文与现代语译及注释的《新编日本古典文学全集·源氏物语》具有参考价值。此外,他指出,尽管国文学研究资料馆、国立国会图书馆等机构已公开大量《源氏物语》数字化资料,学者足不出户即可阅览比对,但即便熟稔草书的研究者,面对《源氏物语》这类鸿篇巨制,也难以精准定位所需内容,若能提供便捷链接,快速访问《校异》或现代语译本等资料,将大大提高研究效率。在此信息共享的基础上,大家一致认为,若能以东京大学本为起点,构建一个可顺畅跳转至《校异》等资料的网站,定将颇有裨益。

4.2 《校异》的文本制作

为了建设学习会讨论中提出的有助于《源氏物语》研究的平台,团队首先着手构建东大本与《校异》之间

的链接。国立国会图书馆数字馆藏已将《校异》以“互联网公开(保护期满)”的形式发布^①,公众可自由查阅。然而,此时尚未发现可供检索的《校异》完整电子文本。因此,中村助理教授下载了所有公开图像并利用 OCR 技术进行处理,从而生成电子文本。尽管 OCR 生成的文本精度超出预期,但仍存在不少识别错误,后续主要由图书馆员承担了人工校对工作。在开展校对工作之前,大家遵循田村副教授的建议,制定了若干文本处理原则,包括:将旧字替换为新字,保留踊字(即叠字符号),以及在文本中插入原文中的换行符。这些规范为后续工作提供了明确的方向,并确保了文本的一致性与完整性。

为便于校对作业,中村助理教授开发了一套专用系统(图 2)。系统左侧显示《校异》的原始画面,右侧则为可编辑状态的未校正文本,方便工作人员在比对后进行实时修改。未校正文本在事前尽可能完成新字与旧字的批量转换,减少手动操作的繁琐。此外,对于反复出现的误识别内容,工作人员随时向中村助理教授反馈,以便对余下的未校正文本进行批量修正,从而有效减轻工作负担。

历经数月的潜心校对,《校异》五卷正文部分的校正悉数完成。秉承原著的编排风格,为文本也插入了相同的换行位置标识,以《校异》独有的页码和行号确保了文本的精准定位。



图 2 《校异》文本的校正系统

另外,《校异》全文文本作为单个独立资源 also 具有很高的利用价值,因此以“校异源氏物语文本数据库”的形式单独公开,采用 CC0 协议作为使用条件^②。

① 芳贺博士纪念会编《校异源氏物语 卷一~卷五》,国立国会图书馆数字馆藏。卷一:<https://dl.ndl.go.jp/pid/3437686>;卷二:<https://dl.ndl.go.jp/pid/3437687>;卷三:<https://dl.ndl.go.jp/pid/3437688>;卷四:<https://dl.ndl.go.jp/pid/3437689>;卷五:<https://dl.ndl.go.jp/pid/3437690>。访问日期:2023 年 10 月 26 日。

② 里源氏学习会“校异源氏物语文本数据库”:<https://kouigenjimonogatari.github.io/>,访问日期:2023 年 10 月 26 日。

4.3 九大本和《校异》之间的链接

随后,学习会着手将《校异》的页码叠加至《源氏物语》抄本的图像上,以建立各版本图像与《校异》之间的关联。学习会上虽已确定以东大本为起点的方针,但鉴于九州大学公开的《源氏物语》古活字版^①与无跋无刊记整版本^②(以下简称“九大本”)的图像上,已预先标注了《源氏物语大成》的页码(该页码与《校异源氏物语》的页码一致)(如图3所示),并且只要注明收藏机构即可自由使用,因此最终决定优先处理九大本,暂时搁置最初计划中优先处理东大本的方案。

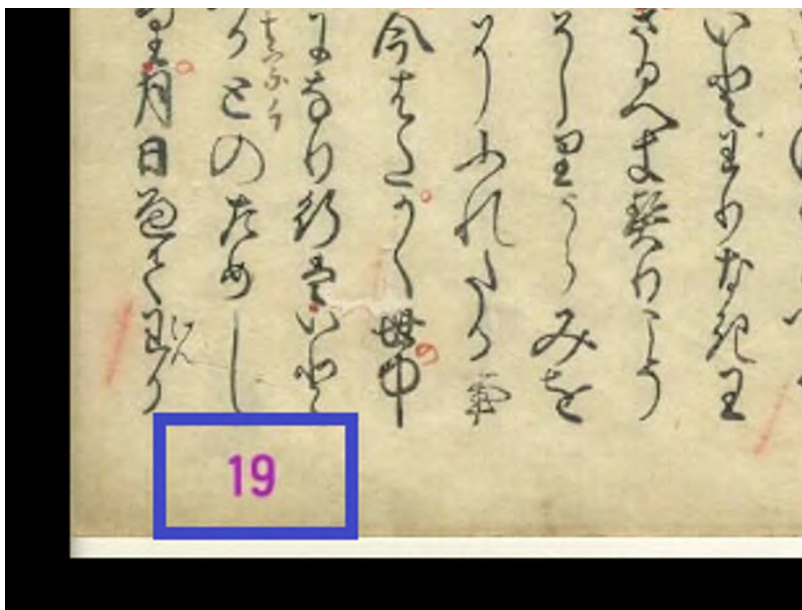


图3 九大本图像中的页码(框内部分)

为顺利开展此项工作,中村助理教授也设计了一套专用系统,主要由图书馆员负责执行。具体而言,图书馆员需对九大本图像进行逐一检查,并记录页码出现位置的坐标。尽管该项工作流程相对简便,但鉴于《校异》全书总计五卷,总页数约达2070页,工作量实属庞大,预计耗时颇长。为了提升工作效率,中村助理教授重新优化了工作方法,具体调整为:(1)以颜色作为依据,自动识别图像中紫色标记的页码位置;(2)运用OCR技术,将页码数值转化为文本数据;(3)将步骤(1)中页码出现位置的行首部分,设置为步骤(2)中获取的页面的起始位置。此举免除了图书馆员人工核对环节,从而显著缩短了工作时间和工作量。

① 九州大学附属图书馆藏《源氏物语》古活字版1《桐壶》:<http://hdl.handle.net/2324/411193>,访问日期:2023年10月26日。

② 九州大学附属图书馆藏《源氏物语》无跋无刊记整版本1《桐壶》:<http://hdl.handle.net/2324/411265>,访问日期:2023年10月26日。

4.4 东大本和《校异》的链接

在完成九大本的文本处理后,学习会转而开始为东大本图像添加《校异》的页码标注。尽管东大本的图像上不像九大本那样预先标有页码,图书馆员原以为“《校异》每页的行数大致相同,应该可以按等间隔进行标注页码”,“凭借九大本的经验,即便无法完全辨识崩字,亦应能大致确定标注位置”。然而,这种预想过于理想化。实际操作中发现,如果对原文内容不具备一定的阅读和理解能力,准确定位其相应位置殊为不易。

在向中村助理教授反映了这一情况后,他迅速地调整了工作流程。首先,为使不熟悉崩字的人员也能参与工作,中村助理教授利用 ROIS-DS 人文学开放数据共享中心(Center for Open Data in the Humanities, CODH)提供的 KuroNet 崩字识别服务^①,生成了东大本图像的 OCR 文本(图 4)。随后,将生成的 OCR 文本与《校异》各页首行文本进行对比,并通过机器计算了两者间的相似度。最后,系统自动将相似度最高的行与《校异》的页码关联起来,并由人工对结果进行审核确认。此方法的引入,改变了过去依赖逐页手动标注页码的困境,工作人员只需在汇总了《校异》文本和 OCR 相似度匹配度较高的东大本文本数据的电子表格上核查相似度较低而疑似不一致的部分。这一改进使作业进展顺利,并大幅度地缩短了所需的工作时间。

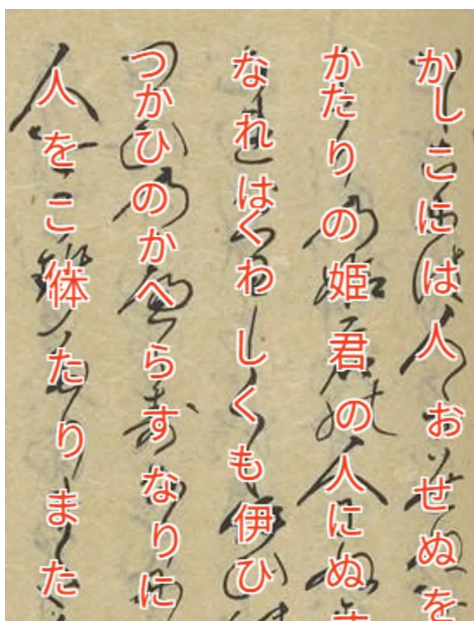


图 4 KuroNet 崩字 OCR 识别的结果画面

^① ROIS-DS 人文开放数据共同利用中心“KuroNet 崩字识别服务 (AI OCR)”:<http://codh.rois.ac.jp/kuronet/>, 访问日期:2023 年 10 月 26 日。

4.5 与现代日语译文的链接

在构建《校异》与现代语译本的链接时,采取了两种途径。其一,与《新编》进行链接,方式与前述东大本的编号体系相仿,因此得以顺利开展工作。值得一提的是,本校现已可以通过 Japan Knowledge 平台访问《新编》,并且在公开之前,团队已与 Japan Knowledge 的相关负责人确认了此种链接方式可行。同时,考虑到系统用户群并非全部为能使用 Japan Knowledge 的本校成员,故设置了两个链接选项——“通过 Japan Knowledge Lib 查看”和“通过 Japan Knowledge Personal 查看”,以满足不同用户的需求。其二,与《源氏物语》的与谢野晶子译本(以下简称“与谢野译本”)进行链接。选择与谢野译本的原因主要是由于此译本可以在青空文库^①上自由获取。日本青空文库上公开的文本或已结束著作权保护期,或已获得了著作权人的许可,因此可以自由使用。学习会对此深表感谢,并积极利用这些资源。

关于这项工作,中村助理教授也设计了一个专用系统(见图5)。首先,利用青空文库所公开的与谢野译本的 HTML 文件,为其每句话分配 ID,在此基础上创建了一个与《校异》文本并排显示的系统。图书馆员在同时查阅这两者的过程中,将每句译文的 ID 标记在《校异》文本相应部分的开头。由于与谢野译本属于现代语译本,起初预想这项任务可以比较轻松地完成。然而,随着工作的深入,团队发现与谢野晶子丰富的表达和对原文独特的解读使得确定其译文与《校异》的一一对应出乎意料地困难。在田村副教授的建议下,团队参考了宫脇文经先生发布的网站“源氏物语的世界 再编辑版”^②以及涩谷荣一先生发布的网站“源氏物语的世界”^③。以上网站对《校异》文本及与谢野译本的内容进行了全面的归纳,为团队提供了极佳的参考,其内容与团队预期中的成果相符,团队对此深感感激并参考了这些内容。

此外,在现代语译的链接工作中,一些本科生同学因受田村副教授课程启发,对学习会活动产生兴趣,也纷纷主动申请参与进来,并以极大的热情圆满完成了任务。如此意外之人力支持,令笔者倍感欣慰与感激。

通过上述工作流程,九大本、东大本的图像、《校异》文本及现代语译相互贯通的准备工作已然就绪。

① 青空文库藏“与谢野晶子译本《源氏物语》”:<https://www.aozora.gr.jp/cards/000052/card5016.html>,访问日期:2023年11月26日。

② 宫脇文经“源氏物语的世界 再编辑版”网站:<http://www.genji-monogatari.net/>,访问日期:2023年11月26日。

③ 涩谷荣一“源氏物语的世界”网站:<http://www.sainet.or.jp/~eshibuya/>,访问日期:2023年11月26日。

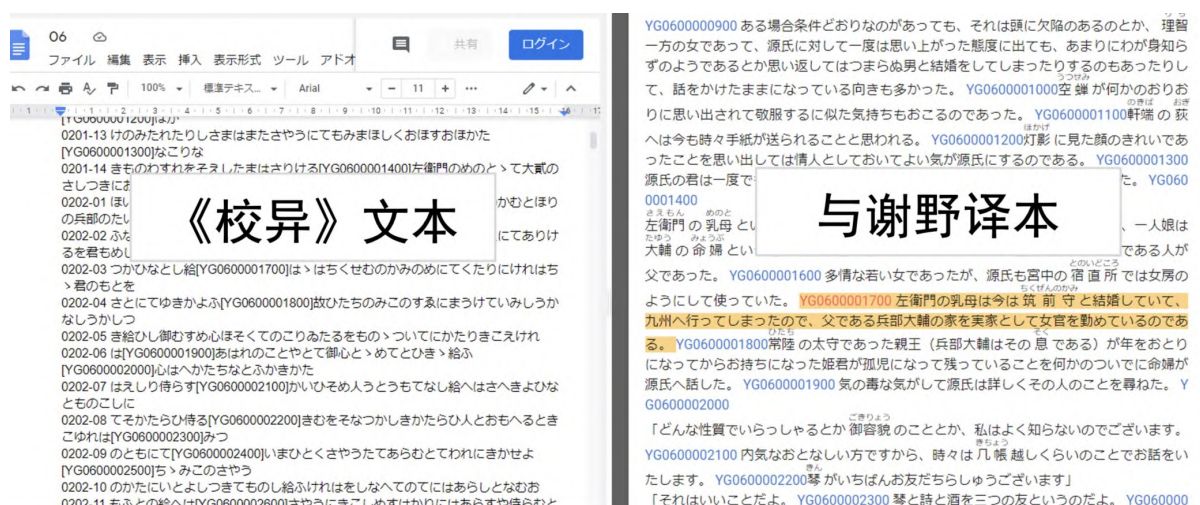


图5 《校异》和与谢野译本的链接工作系统

4.6 “数字源氏物语”的发布

在上述工作的推进过程中,田村副教授收到了来自档案馆化推进室的邀请,拜托他在第3届学术资产档案馆化推进室研讨会上(于2019年11月29日举行)^①就日本国文学领域中数字图像的运用及学习会的相关工作进行演讲。为此,学习会将研讨会当天定为成果网站上线的目标日期。

网站的构建由中村助理教授负责。首先制作了一个初步的界面,左侧显示东大本、九大本及《校异》的图像,右侧则显示文本,呈现出便于对比的布局。文本的排列方式则是将《校异》置于上部,下方为与谢野译本。

《源氏物语》抄本的图像与《校异》及《新编》之间的链接是通过在抄本图像上添加不同的标记来实现的。其中与《校异》的链接以橙色标记,《新编》则以蓝色标记,标记形状皆为银杏叶。这一设计灵感来源于本校的“东大标志”^②,亦象征古人将银杏叶置于典籍中以防虫之习俗。读者可通过这些标记一键访问各书的公开网站并跳转至对应内容。

在《校异》文本的展示区域,每页起始处均设置了 IIF 清单图标,点击该图标,右侧图像会自动切换至相应位置。此外,在《校异》文本中还设置了锚形标记,点击此标记,下方与谢野译本的对应部分将被高亮显示,实现同步浏览。此功能即图6所示“图文共览”的基本功能。

① 东京大学学术资产等档案馆委员会事務局《学术资产档案化推进室主办研讨会》: <https://www.lib.u-tokyo.ac.jp/ja/library/contents/archives-top/seminar>, 访问日期:2023年11月26日。

② 《东京大学标志》: https://www.u-tokyo.ac.jp/ja/about/public-relations/b01_05_01.html, 访问日期:2023年11月26日。

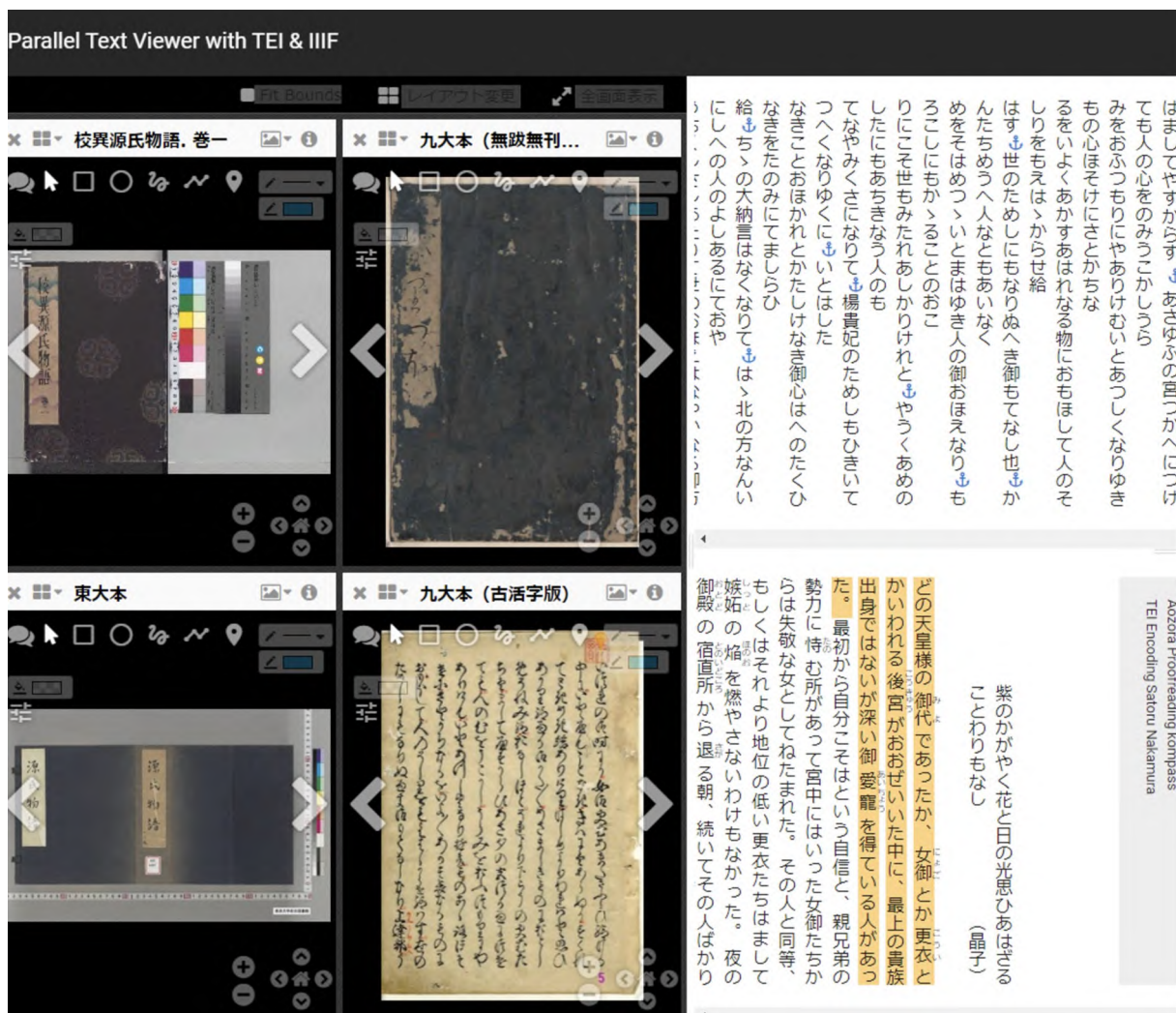


图6 “数字源氏物语”的“图文共览”界面

在研讨会召开之前,虽然项目雏形已调整完毕并具备了公开发布的条件,但考虑到全部数据的整合工作尚未完成,2019年11月29日的首度公开中只发布了其中的《桐壶》卷,标注此次公开版本为“ver.KIRIT-SUBO”也是出于这个原因。

2020年5月,《少女》卷整理完毕,更新为版本“ver.OTOME”;2020年9月,最后一卷最终帖的《梦浮桥》的基本数据整理完成,更新为版本“ver.YUMENOUKIHASHI”。

此外,本系统还提供了以下功能:“页码检索”功能,用户可以一键查找《校异》和《新编》的指定页面;“东大本·脱文错简列表”^①,汇集了工作过程中发现的脱文等信息;“IIIF支持的源氏物语图像列表”^②汇集了当

① 里源氏学会《东大本·脱文错简列表》;<https://genji.dl.itc.u-tokyo.ac.jp/ds/>,访问日期:2023年11月26日。

② 里源氏学会《IIIF对应源氏物语列表》;<https://genji.dl.itc.u-tokyo.ac.jp/list>,访问日期:2023年11月26日。

前已公开且符合 IIIF 标准的近乎完整的《源氏物语》五十四帖图像列表,方便用户浏览。此外,前述的“校异源氏物语文本数据库”也已同步开放。在项目公开后的学习会上,有成员建议增加古注疏著作的链接,以增进实用性,团队遂决定纳入《湖月抄》。《湖月抄》是北村季吟(1625—1705)所撰的《源氏物语》注疏著作。田村副教授指出,该书为《源氏物语》古注疏的代表作之一,不仅收录了原文,还引述了《河海抄》《花鸟余情》《弄花抄》《细流抄》《明星抄》《孟津抄》等其他注疏著作的内容。由于《湖月抄》在江户时代被广泛使用,所以在研究江户文学及近代文学时《湖月抄》具有重要的参考价值。此外,考虑到国文学研究资料馆已将《湖月抄》以自由使用许可(CC BY-SA)的形式公开,遂决定选用该馆收藏的鹤饲文库本^①(索书号 96-20-1~60)进行链接。此链接工作沿用了与东大本及《新编》相同的流程。此外,关于这些工作的详细过程及所涉及的技术细节,敬请读者查阅中村助理教授等撰写的相关论文,以了解更多信息^②。

4.7 工作回顾

先前阐述的《校异》文本校勘与图像文本链接工作,或有繁琐单调之感。然而,通过学习会的交流,团队成员深刻认识到这些工作的意义以及完成后所能为平台实现的功能,从而始终保持着高度的参与热情。此外,得益于中村助理教授精心构建的高效工作系统以及对数据的预处理,团队成员得以在有限的时间内卓有成效地推进工作。由此可见,此次合作并非仅仅停留在专家、系统开发人员和数据整理人员之间简单的分工层面,学习会促成了真正的“协同”工作,其意义不言而喻。

此外,在构建过程中,图书馆员亦提出以下思考。

(1)《校异》作为核心文本,其文本化及赋予其单独的页码和行号对后续数据整理工作产生了深远影响,基础数据构建对项目具有重要意义。

(2)明确标示为可自由使用的图像的便捷性不言而喻。东京大学综合图书馆自 2018 年 6 月起便已将馆藏资料图像在互联网上公开,其使用条件仅为注明馆藏单位名称^③。然而,在实际参与“数字源氏物语”平台建设后,也愈发体会到,使用明确标示为“可自由使用”的图像资源所带来的安心感以及免除申请手续的便捷。虽然或因种种因素而需设定公开数据的使用条件,然而为了最大程度地促进资源的利用,应力求简化使用条件,并将使用条件以通俗易懂的方式予以明示,这一点至关重要。

① 国文学研究资料馆藏《湖月抄》(鹤饲文库);<https://doi.org/10.20730/200018258>,访问日期:2023 年 11 月 26 日。

② 参见中村觉、田村隆、永崎研宣《源氏物语文本研究支持系统“数字源氏物语”开发中 IIIF 和 TEI 的活用》<http://id.nii.ac.jp/1001/00206588/>,访问日期:2023 年 10 月 26 日。

③ 东京大学附属图书馆《图像数据等利用条件》;<https://www.lib.u-tokyo.ac.jp/ja/library/contents/archives-top/reuse>,访问日期:2023 年 11 月 26 日。

(3)通过将《校异》与东大本进行链接的工作,得以发现东大本中存在的脱漏和错简之处,实属难得的体验。作为非中古文学专业的图书馆员,能够有幸参与其中,仿佛窥探到了人文学科研究活动的一隅,令人欣喜之至。

(4)在与现代日语译文进行链接的过程中,通过逐行研读文本,得以领略与谢野晶子笔下所描绘的意蕴丰盈的情感世界,也深切体会到不同译者对原文的理解和阐释存在差异亦属寻常。

(5)人文学科数字化浪潮奔涌,其发展速度远超预想,相关数据与工具也都以开放获取的形态进行公开。

如上所述,团队不仅圆满完成了“数字源氏物语”的构建和公开工作,更使众多图书馆员获益匪浅,深受启迪。

诚如下一节所述,2019年11月“数字源氏物语”发布之后,学习会仍持续开展。然而自2020年春季起,因新冠肺炎疫情蔓延,学习会不得不转为线上形式开展活动。直至2023年5月新冠病毒感染症被日本政府调整为第五类感染症,学习会始终以线上方式保持定期研讨学习,团队得以持续深入地开展各方面的探讨。

5 “数字源氏物语”的功能扩展

“数字源氏物语”的公开版本已迭代至 ver.YUMENOUKIHASHI,构建了一个集成《源氏物语》(3种抄本)、《校异》《新编》及《湖月抄》的图像及《校异》和现代日语译文的文本,实现了图像与文本关联比对功能的平台,这标志着学习会最初设想的“以东大本为起点,无缝跳转至《校异》等相关资料的功能性网站”的建设取得了阶段性成果。因此,学习会继续探讨未来“数字源氏物语”应增设的功能。日本国内尚存诸多基于 IIF 公开的《源氏物语》图像资源,于是大家纷纷提议:是否可以将其余的公开图像资源也纳入平台?是否可以利用数字技术,探索除文本比对以外的功能,以进一步促进《源氏物语》研究?

继而,团队又追加了如下功能:AI 图像检索^①,《绘入源氏物语》的插图比较^②,《源氏百人一首》翻页式面部比较^③。

5.1 AI 图像检索

网络上公开着众多机构所藏之《源氏物语》及其相关资料。虽然期冀“数字源氏物语”平台未来能实现更多版本的图像比对,然而鉴于最终仍需人工对应操作,传统方法工作量实为庞大,在实际操作中面临诸多

① “数字源氏物语”(AI 图像搜索版):<https://genji-ai.web.app/>, 访问日期:2023年11月27日。

② 《绘源氏物语》的插图对比:<https://genji.dl.itc.u-tokyo.ac.jp/picture/eiri>, 访问日期:2023年11月27日。

③ 《源氏百人一首》(翻页式面部比较):<https://genji.dl.itc.u-tokyo.ac.jp/picture/face>, 访问日期:2023年11月27日。

挑战。

如 3.4 节所述,在将《校异》文本行首部分与图像进行关联的工作中,团队进行了预处理,即运用机器自动比对《校异》文本数据与崩字 OCR 识别所得数据,从而遴选出相似度较高的内容。田村副教授据此提出,是否可将此机制应用于所需图像的自动推测。基于此构想,“AI 图像检索”功能应运而生。

“AI 图像检索”功能的实现流程如图 7 所示。例如用户可检索“夕顔”或“犬君”等字符。

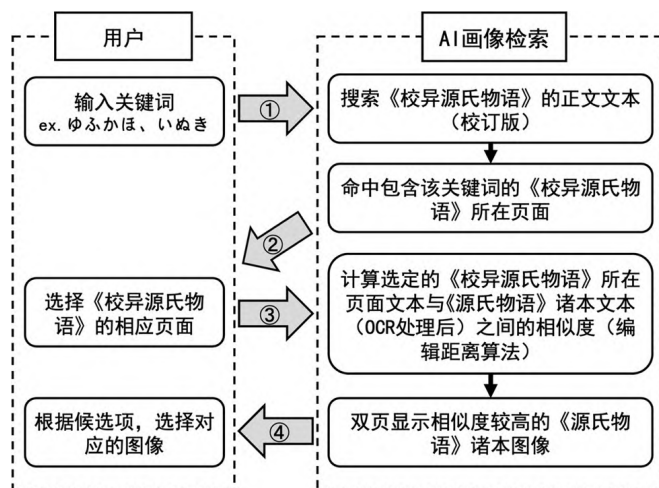


图7 “AI 图像检索”功能的流程

首先,用户需以特定关键词检索《校异》文本(①),并于匹配的候选项中甄别出包含目标关键词的《校异》页面(②,③)。其后,借助“AI 图像检索”功能,系统将呈现与所选页面文本相似度较高的各版本页面图像作为检索结果(④)。如此,用户即可便捷地通过关键词跳转至各版本中特定内容所在的图像。需注意,尽管各版本图像已预先进行崩字 OCR 处理,但 OCR 结果尚未经人工校对。此外,由于该功能以页面为单位进行文本比对,故因版本间差异可能导致页面内文本存在出入,难以确保 100% 的准确性。此机制旨在从多个选项中为用户呈现可能性较高的图像,并最终将确认工作交由用户自主完成。关于该功能的系统设计细节,敬请参阅中村助理教授等人的相关论文。

为实现此功能,除构建上述搜索系统外,团队亦进行了前期的人工准备工作。首先,团队对数据以卷为单位进行了整理以便于检索。本功能所用图像多源自国文学研究资料馆的“国书数据库”,尽管该数据库为每一部作品分配了独立的 URL(URL 清单),然而仅凭 URL 难以区分具体卷次。为此,团队在电子表格中创建了对应表,以明确每卷在数字图像中的起始与终止页数。此项工作主要由彼时亦参与现代语译文链接工作的本科生学生承担。随后,以图书馆员为主体的团队成员承担了《源氏物语》书目信息的确认及增补工作。在电子表

格上构建了待收录书目的候选清单,汇总了相关的书目信息(如书名、刊写版本差异、成书年代、形态、数量、索书号等)及许可信息(如许可条件的公开页面 URL、使用条件)。此举源于团队外部研究者的建言,即在将“数字源氏物语”应用于教学场景时,期望学生能够便捷地查阅各资料的书目信息。团队据此,基于整理所得的清单信息,简化了从检索结果中查阅各资料书目信息的流程。

2020 年 12 月,中村助理教授向学习会成员展示了“数字源氏物语 AI 图像检索版”的雏形。经过多次意见交流和改进,该系统于 2021 年 4 月 27 日正式上线(图 8)。该系统基于《校异》文本,能够快速识别并匹配其他抄本中相似的图像,获得了来自源氏物语研究专家和数字人文领域学者的广泛赞誉。大家对该系统能够轻松比较多幅《源氏物语》特定部分图像的功能感到惊叹,并对古典研究与 AI 技术之融合所拓展的研究领域寄予殷切的期盼。

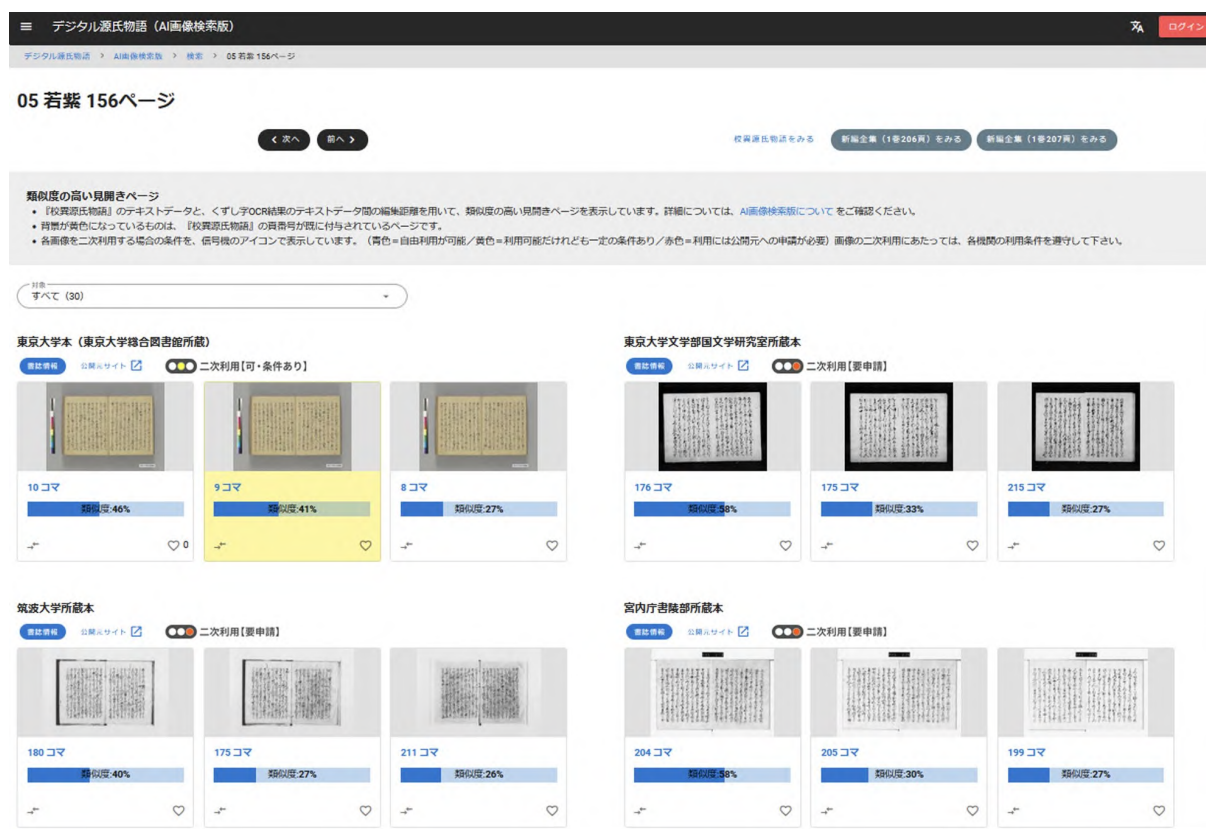


图 8 “数字源氏物语”的 AI 图像检索界面(检索结果)

此外,由于“AI 图像检索”功能能够在无人工干预的情况下导入并自动比对图像,因此未来可便捷地将各机构公开的《源氏物语》图像纳入平台,其重要意义不言而喻。经取得各抄本所藏机构的许可,截至 2023 年 11 月,平台上可供比对的《源氏物语》图像已增至 31 种。

5.2 插图画像的比较

《源氏物语》中的诸多场景,也常以绘画形式展现,例如《源氏物语绘卷》以及各版本中的插图。本节将介绍两个功能,这些功能的实现源于学习会成员提出的一个想法:是否可以通过数字技术对《源氏物语》的插图进行比较?

5.2.1 《绘入源氏物语》插图比较

《绘入源氏物语》是江户时代山本春正编纂的插图版本,文中配有标点符号、浊点、假名注音和旁注,并巧妙地设计了方便读者查阅文中引用的诗歌原典(本歌)的功能。由于这些丰富的附加信息,即使是初学者,也能在无需借助注释书的情况下自行研读《源氏物语》,堪称划时代的书籍^①。《绘入源氏物语》有三种形式:庆安三年(1650年)山本春正所著的“大本”,万治三年(1660年)跋的“横本”,以及无刊记的“小本”(推测刊行于1670年左右)。横本和小本被认为盗版于大本。

鉴于这三种插图版本,学习会成员产生了一个构想:能否将各个版本中相同场景的插图并列呈现?后来也正是这一构想促成了“《绘入源氏物语》插图比较”功能的诞生。借助 IIF 技术自动裁剪已公开的三种《绘入源氏物语》版本中的插图部分,学习会成员对这些图像进行了校对,该功能最终于2021年11月正式上线。

通过此功能,用户能够直观地观察“大本”及其衍生版本“横本”与“小本”的插图在相似性、构图差异以及图像精细程度等方面的异同。宛如于数字空间中设下一张宽广的桌面,便于使用者将三本版本尽情铺陈,以便细致比对^②。作为不同版本文献勘校工具的范例,该功能未来亦有望应用于其他文献资料的比对研究。

此外,田村副教授曾提及《绘入源氏物语》的插图存在重复使用部分元素的现象。在学习过程中,不仅能够从研究人员那里获得关于相关文献的专业知识,还能在电脑屏幕上实际通过观察插图对专业知识进行检验,并从中获得诸多有趣发现,笔者认为这也是受益于学习会促成的图书馆员与研究人员的良好协作。

5.2.2 翻页式面部比较

田村副教授在研究生课程中讲授《源氏百人一首》时,某位研究生偶然发现两个不同版本中人物面部描绘

^① 参见清水妇久子《〈绘源氏物语〉的出版和普及》,高桥亨、久富木原玲、中根千绘编《武士文物与源氏物语绘》(东京:翰林书房,2012年,第343—358页)。

^② 田村隆《〈绘源氏物语〉三种插图比较》:<https://genji.dl.itc.u-tokyo.ac.jp/picture/eiri/about>,访问日期:2023年11月27日。

方式存在差异,由此引发了开发“翻页式面部比较”这款应用程序的想法。这一想法随后被介绍给学习会成员,团队成员由此想到:能否利用数字技术将图像叠加,以便在屏幕上直观地展现这些差异?为了实现这一目标,团队巧妙地运用了其他机构公开的既有技术:借助日本国立国会图书馆(NDL)实验室开发的“图表自动提取程序”^①提取图像,并利用 CODH 开发的“vdiff.js”(差读图像比对服务)^②进行图像对比。最终成功开发了《源氏百人一首》插图对比应用程序,并将其命名为“翻页式面部比较”(“パタパタ顔比較”)。“翻校”是日本的一种传统校对方法(即快速翻动两页来发现差异)，“パタパタ”是其俗称,由于该应用程序能够像“翻校”一样,将两幅图像叠加进行比对,故得此名称。

《源氏百人一首》,系将《源氏物语》中人物与所咏之歌以仿《小仓百人一首》之形式进行介绍的著述。据悉,该书的诸多版本在刊行记录、序言的有无、序言的排列顺序、广告的刊载等方面存在着差异^③。若对诸版本进行细致比对,观察人物之面部的双眼部分,则可发现不同版本中人物双眼之开合状态也存在差异。于“数字源氏物语”中,暂且将二者分别称为“引目”(闭眼)和“开目”(开眼)。

在“翻页式面部比较”功能中,选取了大阪公立大学中百舌鸟图书馆收藏的版本(引目)^④和奈良女子大学学术信息中心收藏的版本(开目)^⑤作为对比对象(图9)。针对特定人物的图像点击应用程序的“对比”按钮后,这两个版本的图像将会重叠,并以不同颜色突出显示差异之处。此外,用户还可以手动拖动重叠图像的分界线,实现类似于校样比对的效果。利用此功能比对图像后,发现图像之差异几乎都限于人物面部,其余部分则完全一致。据此可推测,其中一版本或仅对原版本刻的人物面部进行了重新雕刻并嵌入。此外,进一步细致观察可发现,相比于大阪公立大学中百舌鸟图书馆所藏版本(引目),奈良女子大学学术信息中心所藏版本(开目)图像部分存在重新雕刻的痕迹^⑥。“翻页式面部比较”等数字技术,展现了其在古籍版本对比、印证既有学术成果以及提供丰富信息方面的巨大潜力。“翻页式面部比较”与《绘入源氏物语》的“插画对比”功能一同于2021年11月发布。

① Ndl-lab“图表自动提取程序”;<https://github.com/ndl-lab/tensorflow-deeplab-v3-plus>, 访问日期:2023年11月27日。

② ROIS-DS 人文开放数据共同利用中心“差读(Differential Reading)图像比对服务”;<http://codh.rois.ac.jp/differential-reading/>, 访问日期:2023年11月27日。

③ 参见管宗次《解说》,黑沢翁满编《源氏百人一首》(东京:和泉书院,1999年,第161—167页,和泉书院影印丛刊,91)。

④ 国文学研究资料馆国书数据库,大阪公立大学中百舌鸟图书馆藏《源氏百人一首》;<https://doi.org/10.20730/100097415>, 访问日期:2023年11月27日。

⑤ 国文学研究资料馆国书数据库,奈良女子大学学术信息中心藏《源氏百人一首》;<https://doi.org/10.20730/100241606>, 访问日期:2023年11月27日。

⑥ 例如,在“翻页式面部比较”的第8格中,可以看到“指喰女”脖子上的嵌入痕迹。



图9 “翻页式面部比较”示例

5.3 说明视频的制作

随着“数字源氏物语”功能的日益丰富,清晰地向用户阐释所提供功能的必要性也日益凸显。项目以《源氏物语》文本比对为主体,同时兼具“AI 图像检索”和“插图比对”等性质有所差异的扩展功能。尤其是这些扩展功能,运用了大量诸如“崩字 OCR 识别”和基于相似度(编辑距离)的文本比对等较为复杂的最新技术,因此,初次接触项目的使用者或难以直观理解其功能用途。因此,团队决定制作一段关于“图文共览”“AI 图像检索”“翻页式面部比较”这三种功能的帮助视频。视频主要由学习会中的图书馆职员负责制作。他们以非源氏物语专门研究者的视角,力求呈现一部让同样为非专业研究者的用户亦能轻松理解的视频。团队成员中有一位具备视频制作经验的图书馆员,因此由其牵头负责该项工作。首先,团队成员实际操作了各个功能,并从

中梳理出可能令用户感到困惑的环节,随后以此为基础制作了分镜脚本。接着,根据分镜脚本,利用 PowerPoint 制作了视频所需的幻灯片,并在部分幻灯片中嵌入了实际操作“数字源氏物语”的演示视频。最后,利用 PowerPoint 的视频导出功能将幻灯片转换为视频,并上传至 YouTube 平台^①。

“数字源氏物语”功能繁多,部分内容难以用文字详尽阐述,而视频则可通过结合实际操作更直观地向用户呈现。在如今视频已成为主要信息获取手段的时代,这一说明视频的制作经验不仅提升了参与制作的图书馆职员的专业素养,也对日常的图书馆工作产生了正面影响。

此外,CODH 网站不仅以“翻页式面部比较”为例介绍了 vdiff.js 工具的实际应用,并提供了该说明视频的链接。制作之初,团队并未预料到此视频能够超越“数字源氏物语”而获得其他机构的应用,实在是意外之喜,不胜感激。

5.4 向其他机构申请使用许可

“数字源氏物语”构建初期,团队优先选用遵循例如 CCBY 等许可协议的图像和文本资料,这些素材无需申请即可自由使用。随着“AI 图像检索”功能的上线,项目实现了几乎无需人工干预即可进行的多版本对比,但也由此带来了新的挑战:如何将使用条件限制的图像亦纳入其中,以进一步丰富对比样本?

对于图像的二次利用,有时需要提出申请,有时则存在利用条件未明确标示的情况。为继续丰富“AI 图像搜索检索”功能中收录的版本,团队首先联系了“IIIF 支持的源氏物语列表”中所列举的各收藏机构,并请求其许可,以将其馆藏图像纳入本功能之中。然而此类二次使用与传统意义上刊物中的图像引用有所不同,属于图像搜索工具中的一种特殊应用方式,因此,团队需先向各机构详细阐述“数字源氏物语”及“AI 图像检索”的具体内容。尤其是关于利用崩字 OCR 生成识别文本数据这一环节,在沟通方面存在一定难度,部分机构在做出回应前花费了较长时间,但最终,团队仍获得了多家机构的使用许可。未来,团队也将持续向公开《源氏物语》全 54 帖图像的机构提出申请。

在此过程中,团队深刻体会到随着数字档案功能的日益完善、丰富,未来对于图像资源的二次使用申请形式将呈现出更加多元化的趋势。一些现阶段尚属罕见的应用功能,数年后或许会成为常态。对于此次未能获得许可的机构,团队计划在适当时间后再次与其沟通,并重新提交申请。

5.5 二次使用许可标识的制作

如前所述,由于部分图像资源的使用需经原收藏机构许可,因此,“数字源氏物语”需要以更加直观的方式

^① 说明视频:<https://www.youtube.com/playlist?list=PLJ9pNFDdwPxJ9DCzuRgF1EZfoZAdFkHFo>,访问时间:2024 年 5 月 23 日。

告知用户相关的使用限制。为此,图书馆员经过反复研讨,设计了一套以交通信号灯为创意原型的标识系统。根据使用条件的不同,标识分为三种类型:“绿灯=可自由使用”“黄灯=使用受限,需遵循特定条件”“红灯=需申请使用”。用户无需阅读冗长的文字说明,通过直观的视觉符号即可了解使用限制。此外,每张图像均附有链接,指向其原收藏机构的网站,用户可通过链接查阅详细的使用许可信息。

因此,2022年7月,“数字源氏物语”的“AI图像检索”功能发布了更新版本,除了新增上述二次使用许可标识外,还提供了指向国立国会图书馆藏《校异》版本和 Japan Knowledge 平台收录的《新编》版本对应页面的链接。

6 结语:“里源氏学习会”和“数字源氏物语”的意义

2020年以来,受新冠疫情的影响,人们对数字化数据的关注度空前高涨,数字人文、文本编码倡议(TEI)以及文本利用等领域在日本国内外也日益受到重视。“数字源氏物语”的发布及功能扩展恰逢其时,引起了众多研究者和相关人士的关注与兴趣。项目推广的机会也随之增多,学习会成员田村副教授、中村助理教授、永崎主席研究员等,纷纷结合自身专业领域,在学术会议及期刊上发表了相关论文。图书馆员亦借此机会,在国文学研究资料馆主办的国际研讨会、校内课程及研讨会等场合分享项目成果^①。

回顾本项目的历程,最初本馆馆藏资料的数字化引发了团队的思考以及兴趣:如何超越单纯的“纸本”到“电子数据”的数字化转换,探索更具数字特色的应用模式。最终,运用前沿数字技术,使此项工作成果得以一种前所未有的研究平台的形式呈现于大众之前。笔者深切地体会到,此项成果的取得,很大程度上归功于与田村副教授、中村助理教授等国文学及数字人文领域专家学者的密切协作。例如,围绕着以崩字OCR识别技术识别出的文本是否在未经校对的情况下公开,以及若要公开这些文本该如何避免误导用户等存疑之处,团队进行了多次深入的探讨与意见交流,最终决定开发出平台目前的诸多功能。此外,尽管目前尚未完全实现,但团队也在筹划未来增加英译本等相关资料,并继续探讨如何运用先进的数字技术提升检索精准度。研究人员与图书馆员优势互补,集思广益,方得以构建兼具双方视角的研究平台。或许,这正是“数字源氏物语”的一大特色所在。

如3.1节所述,该项目最初以非官方的学习会形式发起,迄今也未改变初衷。在“数字源氏物语”发布之际,曾有人提出或许应该将活动官方化、正式化,即从“里”转为“表”,以维持平台持续运营。然而,现行非正

^① 田村隆、中村美里、中村觉、永崎研宣《“数字源氏物语”的构建和发展》(第6届日语历史典籍国际研究大会,国文学研究资料馆,2020年),<https://www.nijl.ac.jp/pages/cijproject/sympo2020.html>,访问日期:2023年11月27日。

式的“里”模式亦有其益处:参与者不受工作时间限制,参与门槛较低;校外研究人员以及他馆的图书馆员亦可灵活参与其中。鉴于此,团队决定暂时沿用“里”模式。2022年以来,一些来自其他图书馆的对该项目感兴趣的年轻馆员也积极参与进来,这种跨越组织及部门的互动交流,正是“里”学习会的显著优势。尽管如此,鉴于活动已持续四年,核心成员未来可能陆续离开东京大学,或许有必要重新审视学习会以及“数字源氏物语”的定位。

最后,笔者愿就图书馆员参与项目的意义略述一二。如前所述,这样先进的基于数字技术的研究平台得以建成,实际上很大程度上仰赖于团队成员中专业研究人员的贡献。然而,图书馆员虽然并非是源氏物语研究专家学者,却能从用户的角度出发,对新增功能进行审查,并借鉴在图书馆工作中积累的经验来创建数据、制作解说视频,以及提出与版权相关的建议。此外,在向其他机构申请图像二次使用许可方面,也得益于图书馆之间的交流网络,而能顺利推进。2023年1月,日本文部科学省发布《开放科学时代大学图书馆之应有形态(审议总结)》^①。其中建议:“应以用户为中心,对支持数字化内容利用的各项服务与现有业务进行重新构建。此外,通过大学图书馆之间或与其他学术信息提供机构的合作,促进日本国内学术信息的积累、数字化和流通”。“数字源氏物语”从以研究者为中心的用户视角出发,并利用其他机构公开的图像与工具整合了一个全新的研究平台。此项尝试,可谓契合上述提议。

笔者参与构建“数字源氏物语”,亲身体验近年来数字技术的日新月异。即使不进行本文所述的各种功能开发,信息技术的发展也愈来愈使得基于数字图像开展多种数字人文活动变得可能^②。数字档案馆不应止步于图像公开,更应利用数字技术拓展其附加功能之可能性,随着技术的进步,此类功能的实现也不再遥不可及。冀望“数字源氏物语”之实践经验,亦能为其他机构探索数字档案馆之建设与应用方法提供些许借鉴。

致谢:本文系笔者与田村副教授、中村助理教授、永崎首席研究员等“里源氏学习会”成员及相关人士经共同探讨、实践所得之项目成果。谨向各位给予的鼎力协助表示衷心的感谢。此外,亦衷心感谢为“数字源氏物语”之构建慷慨提供不可或缺的图像数据、文本数据及各项工具等资源的机构,以及慷慨允准利用相关资源的机构。

补充说明(2024年5月27日):关于“3.2.1 东京大学数字档案馆构建项目”一节中所述系统概要,已于2024年5月29日全面更新,系统构成已大幅度更新。敬请查阅该项目网站以获取最新信息。

^① 科学技术·学术审议会·信息委员会·开放科学时代大学图书馆之应有形态审议委员会《开放科学时代大学图书馆之应有形态(审议总结)》:https://www.mext.go.jp/b-menu/shingi/gijyutu/gijyutu29/004/mext_00001.html,访问日期:2023年11月27日。

^② 东京大学综合图书馆对部分内容采用了结合数字档案馆和数字技术的公开方式。例如,综合图书馆所藏的歌德书信及其清稿本《歌德全集》的文本根据 TEI 标准创建了文本数据,并可以进行比较,见“歌德亲笔信函[Letter] Weimar, 29.Dcbr.1822”:<https://iif.dlitc.u-tokyo.ac.jp/repo/s/goethe/page/home>,访问日期:2023年11月27日。

参考文献

[1]秋山虔.源氏物語[M/OL]//日本大百科全書(ニッポニカ).东京:小学館,1994.[2023-10-25].<https://japanknowledge.com/lib/display/?lid=1001000081597>.

Making Use of Digital Data to Build a Research Platform:Developing “Digital Tale of Genji” by Cooperation of Researchers and Library Staff

Nakamura Misa Kigoshi Michi Ogawa Kayoko Trans.Wang Zirui

Abstract The University of Tokyo Library, prompted by the digitization of its collection of The Tale of Genji manuscripts, organized study sessions involving researchers from various disciplines and library staff to explore the utilization of digital data centered around The Tale of Genji. As a result, the website Digital Tale of Genji was launched with the aim of creating a platform for textual research on The Tale of Genji. This paper discusses the background leading to the organization of these study sessions, the development and functional expansion of the Digital Tale of Genji platform, and the significance of collaboration between researchers and library staff in this initiative.

Key words digital archives; digital humanities; humanities and informatics; IIR; faculty-librarian collaboration; The Tale of Genji