

OCR 要点说明



Gale 将会怎样优化“社会中的残疾人”的文字内容，让各种能力水平的用户都能访问使用？

使用屏幕阅读器访问数字化原始文献的研究者依赖准确的文字输出，但因为 OCR 技术的限制，100% 的准确性几乎是不可能的。为了优化“社会中的残疾人”数字档案库的访问性能，Gale 将使用“双 OCR 比较和手工纠正”的方法，确保该档案库中 OCR 文本的准确性至少达到 98%。

这究竟意味着什么？

Gale 将会通过两个 OCR 引擎转录文本：Abbyy 和 Google vision。然后我们会比较两个引擎得到的结果，在二者出现分歧的时候进行校正。

Abbyy 和 Google vision 难于读取的文字怎么办？

在制作过程中会识别出质量较差的图像，然后进行特殊的人工清理过程。Gale 制作团队也会对随机选择的页面进行人工校对步骤，确保转录文字的准确性高于 98%。

“社会中的残疾人”中的所有文字都会经过这种处理吗？

这一处理过程不会应用于手写文本或非英语文本。

图片呢？都会添加替换文字标签吗？

构成这个档案库的图书和小册子包含成千上万张图片，包括插图、图表和照片等。他们不会被添加替换文字，但会被单独划分出来，用户可以在文档导航面板的“插图作品”一栏中查看每张图片。很多图片会配有图片标题，也会经过和 OCR 一样的手工质量保证程序。虽然图片标题并不能取代替换文字，但它们能够提供页面上插图的上下文，可以作为一项功能特色通过辅助技术来访问。

内容如上所述，那么平台本身又会怎样呢？

“社会中的残疾人”中的内容将会在 Gale 原始档案（Gale Primary Sources）平台上提供访问，这个平台符合《Web 内容无障碍指南 2.1》的 AA 级标准。完整详情，包括免责条款，可在 Gale 原始档案 VPAT 中查看：<https://assets.cengage.com/gale/vpat/gdc.pdf>

该档案库发布后，我可以在哪里提供我的反馈意见？

Gale 致力于不断改进我们产品的无障碍功能。我们将在“残疾史”平台的“关于该档案”部分重申这一承诺并提供联系信息。如果您对产品使用体验和辅助技术有任何的问题或反馈，请与我们联系。