

# 手写文本识别[HTR] 技术

手写文本识别(Handwritten Text Recognition, HTR)是一项 革命性的新技术,能够阅读并转录手写文稿。

Gale已经将HTR技术整合到了"犯罪、惩罚和流行文化,1790-1920年"档案库中,对用户而言没有额外的费用,并且这项技术未来也将用在Gale的部分新产品中。

#### 过去是怎样的?

许多年来,由光学字符识别(Optical Character Recognition,OCR)技术生成的机器转录文本已经让打印文稿的全文检索成为可能。由于手写文字字体存在更大的变化,OCR技术无法应用于手写文稿。

"犯罪、惩罚和流行文化"中的手写文献资源之前也是可发现的,但仅仅基于它们的元数据(概要信息)。打印文稿被全文检索时,手写文稿仅有元数据被检索。因此,很多相关文献可能没有出现在用户的检索结果中。

#### 现在有什么不同?

通过Gale创新性的手写文本识别技术, "犯罪、惩罚和流行文化"中的所有手写文稿的全文在用户检索前就已经被数字化转录,也就是说用户从一开始就可以检索该档案库中所有文献的全文。

shout the laker and of May lack he this of thoused at the Phico Bound without the things the changed with Misdomeonous, when the hand how Duncan was so guined to out water his Saint for the said alonged them this tennable bound, and this Depose to separate to such and this Depose to appear and give in down against that the Duncan on the said The Marken Depose in down against the When Depose in the said The last the Depose to the said The last the Depose to the said the Separate of the Said The last the Depose to the said the Secretary and paper delisaid to this the said to the said to this the said to this the said to the said to this the said to the said to this the said to the said to

whomas charged this day before Me, for that wither the Survivadicher of a and 1866 Central Crime Good the distribution of factorist the died from the Control Crime of the distribution of the formal the Secretary of thate predictions of the the Mar IR seem of the Ba Balas and with intend to the man 1882 also died unlampfully & means of firsten pretences to the factorist may be stored for Mar IR seem of fifty will be married prelimines to the false and with intendent to the secretary of the for Mar IR seem of false pretences also died unlampfully by means of false pretences also died unlampfully by means of false pretences also died unlampfully by means of false pretences also died on 1882 to means of false pretences also died on of the pretences to be false and with when the format the transfer of the Mar IR among pretences to be palse and with mideal to the the means of the pretences of the palse and with mideal to the the died on 90° april 1884 to means of palse pretences and died on 90° april 1884 to means of palse pretences to be palse and with intend to the died on 90° april 1884 to means of palse pretences to be palse and with intend to the died on 90° april 1884 to means of palse pretences to be palse and with an about to the died on 90° april 1884 to means of palse palses.

Circular for funds to rebuild

Atte Chancel & organ Chamber

Atta Parish Church of St.

Nicholas.

A fraintet

Circular, of which a spring

more schools were, was have

ap, the contents of which were

discussed between me, Mr.

Rumbald, and Defendant, before

# GALE的手写文本识别技术 打开了一次文献档案的 更多可能性

#### 对用户和图书馆的好处有哪些?

#### 提高检索效率

用户将会从一开始的检索中就得到大量相关检索结果。这将更快地引导研 究者找到相关资料, 节省宝贵的研究时间, 并且将会打开更多可能的研究 路线。

## 没有古文字技能的人 也能访问

手写文稿有时被认为对于本科生而言难度过大, 他们往往缺乏古文字技 能,对手写文稿望而却步。随着整个社会都对手写文字越来越陌生,这个 问题将日趋严重。利用HTR技术得到的转录文本呈现出文献内容,帮助学生 获得对文献的全面理解。此外,它将学生直接指向更为相关的资料,减少 学生们需要处理的文献数量, 让利用手写文稿开展研究变得不那么令人生

# 向更多用户打开历史 档案的大门

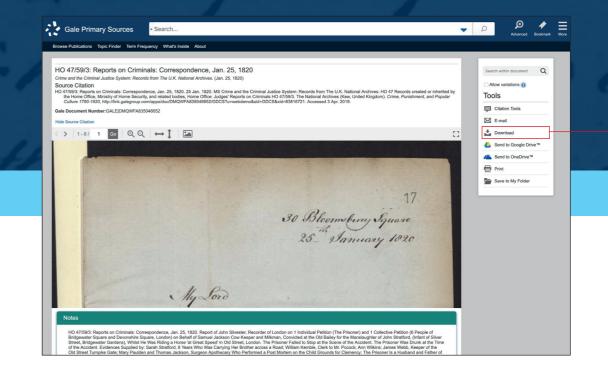
提高这些档案资源的可访问性意味着院校中的潜在用户群变得更大。手稿 不再是只有某些院系使用的小众资源。

### 浏览大量文献变得更 加轻松

即便具备古文字技能的研究者也会发现可轻松阅读的转录文本能够帮助他 们快速浏览一篇文献的内容,让他们能够在固定的时间内查看更多资源。

# 结合到数字人文研究 项目中

之前,在将手写文稿加入到数字人文研究项目之前,需要手工转录这些手 稿。这将花费大量的时间,因此是在数字人文研究项目中使用手稿典藏的 主要障碍。Gale的HTR技术能够自动转录大量的手写文稿。因此,手稿现在 可以在数字人文研究中发挥更大的作用。以这种方式使用"犯罪、惩罚和 流行文化"中的手稿将在Gale数字学术实验室(Gale Digital Scholar Lab) 中特别有效。



通过"下载" (Download) 按钮可以轻松获 取手写文稿的数字转录 文本。

手写文稿的数字转录文本可以与手稿进行对照分析,为学生说明 文献内容, 让资深研究者更轻松快速地处理文献资料, 甚至能够 帮助培养学生们的古文字技能。



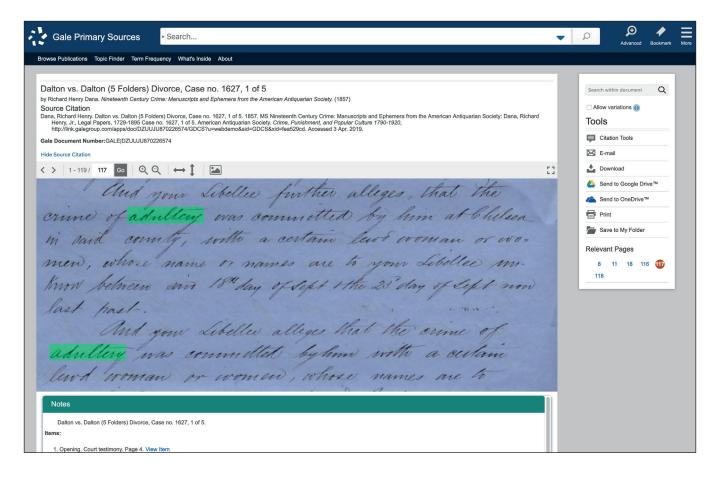
GALE\_DMQWFA835046652.txt

Disclaimer: This file is generated from OCR (optical character recognition), which is a technology that converts images of text into text. While the technology is good at deciphering legible text, there are limitations and some text may not have been extracted correctly, HO 47/59/3: Reports on Criminals: Correspondence, Jan. 25, 1820 in 30 Bloomsbury Square 25 January 1820. My Lord In pursance of the Command of His Royal Highness the Rince Regent communicated to me by your Lordships Letter of 21. Iorstant I have sent enclosed herewith a Report of the Case of Samuel Jackson which I have perused and fully considered, and think the Prisoner hath been guilty of very great and Culpable Nequence, that the sentence is a very lenient one and such as I cannot recommend to His Royal Highnoss for a Miligation of Jam with Respect My Lord Your Lordshiops Most Obedient Humble servant John Silvester Lord discount Sidmenk. GALE DMQWFA835046652.tx quad it 30 Bloomsban 25 empty 1820- the Recorder Report on the Case. of

Sidmenk.
squad it 30 Bloomsban 25 empty 1820— the Recorder Report on the Case. of cancer sacksion
19 Samuel Jackson was Indicted for Feloniously Hilling and slaying John Stratford. Sarah Stratford said she was light your of age that on 12. November she was carrying her little brother up Old Street and as she was crossing the Road the Prisoner came on Horseback full Gallop, and the store knocked her down forced he brother out of her arms, the continued riding on as fast as before. habrothers eye was knocked out t he was bloody but was alwe when she took him up The peopleteck him to doctor Jackson and he lived three days. Upon being less Cxammed she saw s he did not see the Horse coming that she was fughtened when it came up; The said she did not drop her brother but the Horse knocked him out of her Arms; and that herfather did not tell her what to say, William Hemble said he was Glerk to Mr Brock On 12. Nov being in the Street he saw the Prisoner, whow then a bout forty yards from the Tunpikeljate on horseback riding along full Gallop, without a saddle, but he believed he had a he passed through the Sinpike at the same rate without pay. He said he was about Two hundred yards fomthe Prisoner when first saw him The little Girl with the Infant in her Arms was crossing and had just passed the lochdljate when the Horse struck her and forced the Infant out of her are that he called to the Prisoner to stop before he passed him and also after the Prisoner had looked round and must have seen the Children on the the Ground, but he continued galloping on the he knew he was called to and pursued, but finding an alarm was given knowing he must be stopped he at last pulled up: that he went up to him and laid hold of the Bridle I another person common up who was an Officer he gave him charge of the Privoner who got off the Howe, that he told him he had rode over a Child, which he treated with the greatest indifference and said bont make a fuss it cannot be helped. He said the Prisoner was very much intoxicated. He admitted tha

lso said the Pcidone had an opportunity of stopping the horse for on inding he was pursued and that it would be useless to side further he finding he was pursued and that it would be useless to side further he stopped the horse in a moment. Upon being asked by the Court If the Frisoner when he passed him appeared to be trying to stop the Horse or encor= raging it to goon. He said encouraging it to go on. Ann Wilkins said she was in On Street Road sitting by the Turnpike on the other side of the way and heard a Horse coming violently along and as soon as it got through the Gate she heard somebody try Htattoe, she looked round saw two Children laying on the Ground and the man the House were close to them the man struck his Horse with a this or stick and rode off as fast as he

Gale创建手写文稿转录文本时没有筛选或挑选过程,因此Gale的HTR技术捕获了所有的手写文本,包括拼写 错误、古代拼写方式和错误识别的单词,让研究者能够自行确定哪些是真正相关的。



之前,如果用户检索"adultery"(通奸)一词,这篇文献不会出现在检索结果中,因为这个词并没有在元数据中使用。然而,HTR技术显示出"adultery"一词在这篇文献的多个地方出现。因为现在用户一开始检索时就能够检索文献的全文,这篇一次文献将会出现在他们的检索结果中。

手写文本识别应用于"犯罪、惩罚和流行文化,1790-1920年"档案库中,没有任何附加费用,未来也将会应用在Gale的部分新档案产品中。

了解更多信息,请访问: gale.com/cn。

