

ABBYY® Recognition Server



Royal
Botanic Garden
Edinburgh

Das Herbarium des königlichen botanischen Gartens von Edinburgh wird durch ABBYY Recognition Server revolutioniert

Leistungsfähige Server-OCR-Technologie erweitert Datenbank und ermöglicht Online-Suche für wertvolle Sammlung

Die reichhaltige Sammlung lebender Pflanzen des Royal Botanic Garden Edinburgh (RBGE) erstreckt sich über vier Gärten und beheimatet ein weltbekanntes Forschungs- und Lehrzentrum für Pflanzenkunde. Sein Auftrag ist es, die Flora zu erforschen, zu erklären und so für künftige Generationen zu bewahren. Seit über 330 Jahren kultiviert und untersucht dieser altherwürdige botanische Garten Pflanzen und spielt dadurch eine wichtige Rolle für die Dokumentation und den Erhalt der Pflanzenvielfalt unseres Planeten. Vor kurzem begann der RBGE ein Projekt, das derzeit viele akademische Institutionen und Bibliotheken angehen: Die Online-Bereitstellung eines Teils ihrer Sammlung, nicht nur für interne Nutzung sondern auch für die breite Öffentlichkeit.

Die Herausforderung

Da Pflanzen und Ökosysteme heutzutage überall auf der Welt massiv bedroht sind, besteht ein großer Bedarf an Wissen über die Flora, um wirksame Schutz- und Erhaltungsmaßnahmen ergreifen zu können. Dank seiner langen Tradition in der Botanik und Biotopforschung bringt der Royal Botanic Garden Edinburgh beste Voraussetzungen als Dienstleister für Wissenschaftler und engagierte Laien mit. Das umfangreiche Herbarium des botanischen Gartens in Edinburgh stellt die Hälfte bis zwei Drittel der weltweiten Flora dar und gilt damit als eine der wichtigsten botanischen Sammlungen der Welt. Die Besucher der Institution reisen aus der ganzen Welt an, um Einsicht in das Pflanzenmaterial zu nehmen. Der RBGE suchte nach einer Lösung, mit deren Hilfe das Auffinden wichtiger Daten über Pflanzenarten automatisiert und der Zugang zur Sammlung über das Internet erweitert werden könnte.

Dabei sah sich das Institut der enormen Aufgabe gegenüber, ihre umfangreiche Sammlung von drei Millionen Millionen gepressten Pflanzenmustern und Katalogdaten zu digitalisieren. Insbesondere suchten sie nach einer Lösung, die das Erfassen von Text auf Etiketten und Schildchen der Pflanzen sowie auf Papierdokumenten der Sammlung ermöglichen würde. Bislang wurde diese Arbeit überwiegend durch manuelle Dateneingabe eines Mitarbeiters bewerkstelligt – ein überaus zeitaufwändiger Prozess, der zu unvollständige Datensätzen des Forschungsinstituts führte.

Eine zusätzliche Herausforderung lag in der Tatsache, dass die Originalmaterialien unterschiedlichste Formate aufwiesen, mit hand- oder maschinenschriftlicher Information auf Etiketten, die teilweise noch aus dem späten 17. Jahrhundert stammten. Das Digitalisierungsunterfangen wurde überdies verkompliziert durch vielfältige Schriftarten, Sprachen und Barcodes. Angesichts der großen Relevanz dieser Informationen für die Forschung musste die Erfassungsmethode für die Etiketten höchst effizient sein und minimalen Informationsverlust garantieren.

Die Lösung

Als sich der RBGE dazu entschloss, eine Alternativlösung für das Projekt ausfindig zu machen, standen zwei Ziele im Vordergrund: Zum einen musste die neue Lösung Text auf Pflanzenetiketten und Papierdokumenten der Sammlung ohne Informationsverlust erfassen können – selbst von Materialien mit schlechter Qualität oder komplexer Struktur. Zum anderen wurde eine Lösung benötigt, die sich in den bestehenden Workflow eingliedern lassen sowie auf das Bild-Management-System des Instituts, das die TIFFs speicherte, zugreifen können würde.

Über Royal Botanic Garden Edinburgh

Der Royal Botanic Garden Edinburghs (RBGE) wurde im 17. Jahrhundert als Heilkräutergarten gegründet. Heute erstreckt sich die reichhaltige Sammlung lebender Pflanzen über vier Gärten und beheimatet ein weltbekanntes Forschungs- und Lehrzentrum für Pflanzenkunde.

Kontakt

Royal Botanic Garden Edinburgh
20A Inverleith Row
Edinburgh
EH3 5LR
Vereinigtes Königreich
+44 (0) 131 552 7171
www.rbge.org.uk

ABBYY® Recognition Server

Durch den Besuch eines Seminars an der Britischen Nationalbibliothek wurde der RBGE auf OCR (Optical Character Recognition)-Technologie aufmerksam. OCR-Programme können Text aus gescannten Dokumenten oder digitalen Fotos von Papierdokumenten und PDFs zu auslesen und in bearbeitbaren und weiterverwendbaren Text umwandeln. Auf Empfehlung anderer botanischer Forschungseinrichtungen und der Britischen Nationalbibliothek entschied sich das RBGE-Team für ABBYY Recognition Server als ideale Lösung für ihr Digitalisierungsprojekt. Recognition Server ist eine robuste, skalierbare Lösung zur Automatisierung von Erkennung und Umwandlung großer Mengen an Dokumenten und PDFs.

ABBYY Recognition Server bietet Umwandlung von Bildern in Textdokumente mit höchster Genauigkeit zum Zwecke der Klassifizierung, Suche und Ausgabe der Daten an das institutseigene Dokumentenarchivierungs- und Dokumentenmanagement-System. Recognition Server greift auf den Bestand an TIFFs in einem Ordner des Bild-Management-Systems des RBGE zu. Nach der Verarbeitung der hochauflösenden Pflanzenabbildungen erstellt Recognition Server zwei Ausgabedateien: ein durchsuchbares PDF, das als Sicherungskopie gespeichert wird, und eine .txt-Datei unter einem speziellen Ordner des Servers. Die RBGE-interne Workflow-Software holt die Textdatei von diesem Ablageplatz ab und fügt sie einer MySQL-Datenbank hinzu. Über die Datenbank wird die umfassende Institutsammlung für Wissenschaftler leicht nach vielfältigen Kriterien durchsuchbar. Außerdem sind alle Aufzeichnungen der internen Datenbank nun auch online verfügbar.

Das Ergebnis

Durch die Implementierung von ABBYY Recognition Server ergeben sich zahlreiche Vorteile für den RBGE: So verfügt das Institut nun über ein zuverlässiges und leistungsstarkes System und spart drastisch Zeit ein, die zuvor für die manuelle Dateneingabe erforderlich war. Ein weiterer großer Vorteil ist die Verfügbarkeit von umfassenden Informationen über jede einzelne Pflanze in der Datenbank. Das Forschungsinstitut führt dies auf die Fähigkeit des Recognition Server zurück, enorm vielfältige Textmaterialien überaus genau zu erkennen und umzuwandeln – darunter eine Vielzahl an Schriftarten und -größen und zahlreiche Sprachen (47 mit Wörterbuchunterstützung und weitere 139 Erkennungssprachen).

„Die leichte Durchsuchbarkeit unserer Sammlung als Datenbank verdanken wir im Wesentlichen ABBYY Recognition Server. Die schnelle Auffindbarkeit von Pflanzenmustern anhand vielfältiger Etiketteninformationen verschafft unserer akademischen Forschung eine starke Ausgangsposition“, sagt Dr. Elspeth Haston, stellvertretende Kuratorin für Digitalisierung am Forschungsinstitut des Royal Botanic Garden Edinburgh.

Der RBGE ist außerdem sehr zufrieden mit dem erweiterten Zugang zu seinen Sammlungen, den es Forschern innerhalb und außerhalb der Organisation bieten kann. In der Vergangenheit mussten Wissenschaftler Einiges an Geld und Zeit investieren, um entweder nach Schottland zu reisen, um dort die Pflanzen zu studieren, oder um einen Antrag auf Leihgabe an den RBGE zu stellen, die anschließend postalisch versandt wurde. Über den Internetzugang zur Sammlung können Forscher nun überall und rund um die Uhr sofort auf die benötigten Informationen zugreifen.

„Recognition Server hat uns unserem Ziel, die außerordentlich wertvolle RBGE-Sammlung von Millionen Pflanzenbeispielen komplett zu digitalisieren und letztlich der ganzen Welt zugänglich zu machen und so für künftige Generationen zu bewahren, ein gutes Stück näher gebracht. Die Ergebnisse und die Leistungsfähigkeit von ABBYYs OCR-Technologie haben uns restlos überzeugt, und wir beabsichtigen daher ihren Einsatz auch für künftige Digitalisierungsprojekte,“ erklärt Dr. Haston.



Über ABBYY

ABBYY ist ein führendes Unternehmen in der Entwicklung von Technologien für Dokumenterkennung, Dokumentumwandlung, Data Capture und Linguistik. Zum Produktportfolio von ABBYY gehören: **FineReader OCR** und **PDF Transformer** – Endanwenderprogramme zur Umwandlung von Dokumenten; **Recognition Server** – eine serverbasierte Lösung für OCR und PDF-Umwandlung; **FlexiCapture** – Data Capture Lösung zur Verarbeitung von Formularen, semi- und unstrukturierten Dokumenten; **FineReader Engine SDKs** mit dem gesamten Leistungsumfang der ABBYY OCR-Technologien; **Lingvo** – eine Serie von elektronischen Wörterbüchern.

Mehr Informationen über ABBYY unter www.ABBYY.com