Ecco un'altra novità che unisce l'utile al dilettevole. Finalmente, forse per la prima, navigare su internet e usare siti famosissimi come face book etc, aiuterà a salvare la letteratura, i libri scritti. Avete presente quei test in cui, per accedere alla risorsa contenuta in una pagina web o per intervenire in un blog, si deve scrivere una sequenza di parole o numeri che appaiono sfocati e sembrano buttati lì sullo schermo? Sono i Captcha, test inventati per contrastare le sofisticate tecniche degli spammer. Ebbene, quando riportate le parole di questi Captcha, potrebbe darsi che stiate dando un contributo alla digitalizzazione di testi antichi. E quindi alla loro salvezza. Da un anno, infatti, le cronache, le riflessioni, le poesie e i racconti conservati nelle biblioteche hanno milioni di nuovi alleati, che lavorano senza saperlo. Finora, grazie ad un'idea semplice ma innovativa, sono state tradotte in formato digitale 440 milioni di parole, l'equivalente di 17.600 volumi.

I progetti di digitalizzazione dei libri sono moltissimi, perché trasformare l'inchiostro in bit consente di preservare i loro contenuti e di renderli disponibili on line. Di solito si passano le pagine allo scanner e si sottopongono le immagini ad un software di riconoscimento ottico dei caratteri (OCR), che trasforma i testi in un formato riconoscibile dai computer. Il problema è che spesso la carta è ingiallita e le lettere sono poco leggibili, quindi il programma ha bisogno dell'aiuto di un essere umano. E l'intervento di un operatore, ovviamente, costa e rallenta enormemente il procedimento.

Un gruppo di ricercatori della Carnegie Mellon University ha avuto un'intuizione. Ogni giorno, attraverso i Captcha, decine di milioni di persone decifrano su internet delle parole distorte per dimostrare di non essere dei software automatici che cercano di diffondere spam. Perché non sfruttare questa enorme forza lavoro gratuita per dare una mano ai software OCR?

Il successo dell'iniziativa è stato sorprendente. Nei primi dodici mesi i visitatori di circa 40mila siti web hanno decifrato 440 milioni di parole con un'accuratezza del 99%. "Attualmente vengono tradotte 4 milioni di parole al giorno - dice Luis Von Ahn, uno dei ricercatori coinvolti nel progetto - Per ottenere i risultati che raggiungiamo in una settimana, più di 1.500 persone dovrebbero lavorare per 40 ore a testa ad un ritmo di 60 parole al minuto".

ReCaptcha collabora con l'Internet Archive, un'organizzazione non profit che digitalizza i libri di 70 biblioteche e università statunitensi, e con il New York Times, che intende così salvaguardare il suo archivio. Secondo i suoi creatori, che hanno presentato il progetto sull'ultimo numero di Science, questo è solo l'inizio. Attualmente viene tradotto l'equivalente di 160 libri al giorno, ma la diffusione del nuovo sistema potrebbe far lievitare queste cifre, salvando intere biblioteche. Anche lo spam, in un modo o nell'altro, alla fine può rivelarsi utile.

Rezarta Selam Eminaj – DEApress