



The Crowd Optimizes Cost and Quality of Massive Data Project

Crowdsourcing results in 70% reduction in cost and a 30% improvement in accuracy.

The Company

Founded in 1995, eBay Inc. is a global technology corporation. The company manages eBay.com, an online marketplace that connects millions of buyers and sellers worldwide.

The Challenge

Cost-Effective Use of Human Brainpower

eBay features millions of product taxonomies that originate from users all over the globe. The job of ensuring its constantly changing online catalog is optimally organized at all times can't be done with computer algorithms alone. Human brainpower is needed to accurately categorize the products and create product taxonomies from the user's standpoint.

“eBay has completed more than 15 different types of high-value projects through CrowdFlower over the past year.”

In addition to product categorization, eBay needed an efficient method of finding a product's Global Trade Item Number (GTIN). This unique 12- to 14-digit identifier is often missing from posted product descriptions, and the necessary task of finding it is complicated by the lack of a central repository of GTINs.

The company tried to leverage offshore teams in several low-cost destinations to complete these processes, but the traditional outsourcing model posed several challenges with respect to cost, scalability, and accuracy.

The Solution

Microtasking + Automated Quality Control

CrowdFlower's platform took huge amounts of eBay product information and broke it down into microtasks that were completed online by thousands of individuals collaborating across the globe. This massive collaborative effort ensured the enormous data task was accurately completed in record time and at low cost.

With CrowdFlower managing the workflow and checking the accuracy of the data returned, the final results sold eBay on the benefits of crowdsourcing. eBay has completed more than 15 different types of high-value projects through CrowdFlower over the past year.

To improve the product categorization algorithm, eBay and CrowdFlower collaborated to design a machine learning workflow. Through crowdsourcing, they were able to accomplish the job faster, at a lower cost, and with the best possible accuracy level. The machine workflow design, still in use for ongoing classification jobs, presents a contributor with a Product Image and a Product Title with one to six possible classifications.



The Challenge

- Categorize the millions of products in eBay's catalog with accurate and thorough attributes.
- Control costs by scaling outsourcing team to the volume of work required in a given period.

The Solution

- CrowdFlower broke down the project into microtasks for thousands of online contributors.
- A machine learning workflow and other automated processes were used to optimize the data quality.

The Results

- Products were categorized five times faster than with a traditional outsourcing vendor with a 70% reduction in cost and a 30% improvement in accuracy.
- CrowdFlower collected product attributes 10 times faster than a traditional outsourcing firm with a 90% accuracy rate and total cost savings of 60%.
- eBay saved money by reducing the size of the workforce during periods of low demand.

To minimize the impact of first-response bias, the order of the possible classifications is randomized each time it is presented.

In addition, CrowdFlower was able to create a geographic filter to ensure native English speakers who make up eBay's core customer base and hence are more familiar with eBay product taxonomy completed the project.

eBay and CrowdFlower also designed an exhaustive search process for product GTINs. CrowdFlower contributors searched for GTINs for a specific product (based on product title, product type, and image) through a variety of reliable channels. CrowdFlower then compared the contributors' responses, ensuring that each piece of data was verified and accurate.

The Results

The Superior Accuracy of an Iterative Process

CrowdFlower's large online contributor pool completed product categorization five times faster than a traditional outsourcing team with a vastly higher accuracy rate. Global Trade Item Number (GTIN) collection was performed 10 times faster with a comparable accuracy rate.

CrowdFlower's platform collected multiple responses for each product to minimize the impact of any individual error or response. CrowdFlower also increased the number of judgments collected on products that are difficult to categorize or for which it is difficult to find GTINs. This means that when there was disagreement, CrowdFlower collected judgments until a reliable answer was found.

eBay needed a solution that was optimized for both quality and cost and found that solution with CrowdFlower. CrowdFlower provided the quality and accuracy of an in-house team but at a lower cost and in a highly scalable manner.

“As more and more projects with CrowdFlower graduate to production, different groups (engineering, product management, quality engineering) across the organization are seeing the benefits of crowdsourcing and are keen on embracing this new paradigm.”

—Ram Rampalli, Product Management Team for the Selling and Catalogs Group, eBay

About CrowdFlower

CrowdFlower combines human intelligence with the scalability and efficiency of computer algorithms to offer quality-ensured processing of business information. CrowdFlower's platform provides solutions to a wide variety of data needs for enterprises such as product catalog enhancement, content generation, image moderation, and business listing enrichment. With 5 million Crowd Contributors completing millions of judgments each month for over 500 customers, CrowdFlower is the leader in enterprise crowdsourcing. The company has successfully worked with well-respected enterprises including Apple, AT&T, Autodesk, eBay, Ford, LinkedIn, Microsoft, Sears, Toshiba and Twitter.

For more information, visit www.crowdflower.com.



CrowdFlower, Inc. • 2111 Mission Street, Suite 302, San Francisco, CA 94110 • (415) 471-1920

Copyright © 2013 CrowdFlower. All rights reserved. CrowdFlower is a registered trademark in the U.S.A. and certain other countries. All other trademarks or registered trademarks, product names and company names or logos cited are the property of their respective owners.