

With our flagship *DataInFormationSM* suite of solutions for Image Labeling, Data Annotation, NLP Calibration and Form Data Capture, Liberty Source is proud to serve a wide range of clients across many industry verticals. Our *Solution Briefs* provide a snapshot of the challenges clients bring to us, and the outcomes we achieve for them.

Semi-Structured Text Annotation

Industry	Technology	Data Type	PDF
Project Duration	2 Months	Ongoing?	Yes

Challenge

A seemingly simple process of parsing data from a PDF is made increasingly difficult when the format of the document changes frequently. It is time-consuming and expensive to keep the computer model updated to extract the data requested.

Solution

Our Data Associates reviewed the documents and used context clues to extract the desired information from each PDF quarterly report, navigating nuanced different presentations of the information.

Outcome

The resulting data supports tracking consumer trends throughout the current year as well as Year-Over-Year comparisons. This makes various analyses of the data easier, more consistent, and standardized across a variety of different documents.

Semi-Structured Text Annotation

Raw Data from SEC 10-Q form

UNITED STATES
SECURITIES AND EXCHANGE COMMISSION
WASHINGTON, DC 20549

FORM 10-Q
QUARTERLY REPORT PURSUANT TO SECTION 13 or 15(d)
OF THE SECURITIES EXCHANGE ACT OF 1934
FOR THE QUARTER ENDED MARCH 31, 2020

13-0871985
(Commission file number)

INTERNATIONAL BUSINESS MACHINES CORPORATION
(Exact name of registrant as specified in its charter)

New York
(State of incorporation) 13-0871985
(IRS employer identification number)

One New Orchard Road
Armonk, New York
(Address of principal executive offices)

914-499-1900
(Registrant's telephone number)

10504
(Zip Code)

Securities registered pursuant to Section 12(b) of the Act:

Title of each class	Trading symbol	Name of each exchange on which registered
Capital stock, par value \$20 per share	IBM	New York Stock Exchange
2.750% Notes due 2020	IBM 20B	NYSE CME
1.875% Notes due 2020	IBM 20A	New York Stock Exchange
0.500% Notes due 2021	IBM 21B	New York Stock Exchange
2.625% Notes due 2022	IBM 22A	New York Stock Exchange
1.250% Notes due 2023	IBM 23A	New York Stock Exchange
0.375% Notes due 2023	IBM 23B	New York Stock Exchange
1.125% Notes due 2024	IBM 24A	New York Stock Exchange
2.875% Notes due 2025	IBM 25A	New York Stock Exchange
0.875% Notes due 2025	IBM 25B	New York Stock Exchange
0.875% Notes due 2025	IBM 25C	New York Stock Exchange
0.300% Notes due 2026	IBM 26B	New York Stock Exchange
1.250% Notes due 2027	IBM 27B	New York Stock Exchange
0.300% Notes due 2028	IBM 28B	New York Stock Exchange
1.750% Notes due 2028	IBM 28A	New York Stock Exchange
1.500% Notes due 2029	IBM 29	New York Stock Exchange
1.750% Notes due 2031	IBM 31	New York Stock Exchange
0.650% Notes due 2032	IBM 32A	New York Stock Exchange
1.200% Notes due 2040	IBM 40	New York Stock Exchange
7.00% Debentures due 2023	IBM 25	New York Stock Exchange
8.2% Debentures due 2027	IBM 27	New York Stock Exchange
6.50% Debentures due 2028	IBM 28	New York Stock Exchange
7.00% Debentures due 2045	IBM 45	New York Stock Exchange
7.125% Debentures due 2096	IBM 96	New York Stock Exchange

Indicate by check mark whether the registrant (1) has filed all reports required to be filed by Section 13 or 15(d) of the Securities Exchange Act of 1934 during the preceding 12 months (or for such shorter period that the registrant was required to file such reports), and (2) has been subject to such filing requirements for the past 90 days. Yes No

Indicate by check mark whether the registrant has submitted electronically every Interactive Data File required to be submitted pursuant to Rule 405 of Regulation S-T (232.405 of this chapter) during the preceding 12 months (or for such shorter period that the registrant was required to submit such files). Yes No

Indicate by check mark whether the registrant is a large accelerated filer, an accelerated filer, a non-accelerated filer, a smaller reporting company, or an emerging growth company. See the definitions of "large accelerated filer," "accelerated filer," "smaller reporting company," and "emerging growth company" in Rule 120-2 of the Exchange Act.

Data staged for model-assisted annotation

JSON Output

```
{
  'id': '142730',
  'documentId': '9de710d2-1324-4059-9623-7daaf0149f6',
  'bboxLabelClassId': '2438',
  'bboxLabelClassName': 'Company Name',
  'caption': None,
  'shapes': [
    {
      'pageIndex': 1,
      'points': [
        {
          'x': 43, 'y': 141,
          'x': 314, 'y': 141,
          'x': 314, 'y': 176,
          'x': 43, 'y': 176
        ]
      ],
      'status': 'ACCEPTED',
      'labeledBy': 'AUTO',
      'labeledByUserId': None,
      'acceptedByUserId': None,
      'rejectedByUserId': None
    }
  ],
  'id': '142577',
  'documentId': '9de710d2-1324-4059-9623-7daaf0149f6',
  'bboxLabelClassId': '2438',
  'bboxLabelClassName': 'Company Name',
  'caption': None,
  'shapes': [
    {
      'pageIndex': 0,
      'points': [
        {
          'x': 40, 'y': 138,
          'x': 305, 'y': 138,
          'x': 305, 'y': 172,
          'x': 40, 'y': 172
        ]
      ],
      'status': 'ACCEPTED',
      'labeledBy': 'AUTO',
      'labeledByUserId': None,
      'acceptedByUserId': None,
      'rejectedByUserId': None
    }
  ]
}
```



DataInFormation annotators upload 'first pass' annotated data into our database to verify the accuracy of the appended labels and focus on improving the efficiency and quality of the automated annotation, while ensuring the utmost data security