

With our flagship *DataInFormation<sup>SM</sup>* 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.

### Corporate Email Content Mining

Industry	Capital Markets	Data Type	Semi-Structured (.pdf, .txt, .html)
Project Duration	2 Months	Ongoing?	Yes

### Challenge

High volumes of documents pertaining to past, current, and future client and prospect activities are received daily and require parsing and annotation in order to be analyzed for patterns and potential action items. With a very high level of variability among the documents, accurately identifying relevant text strings is not possible in a completely automated environment.

### Solution

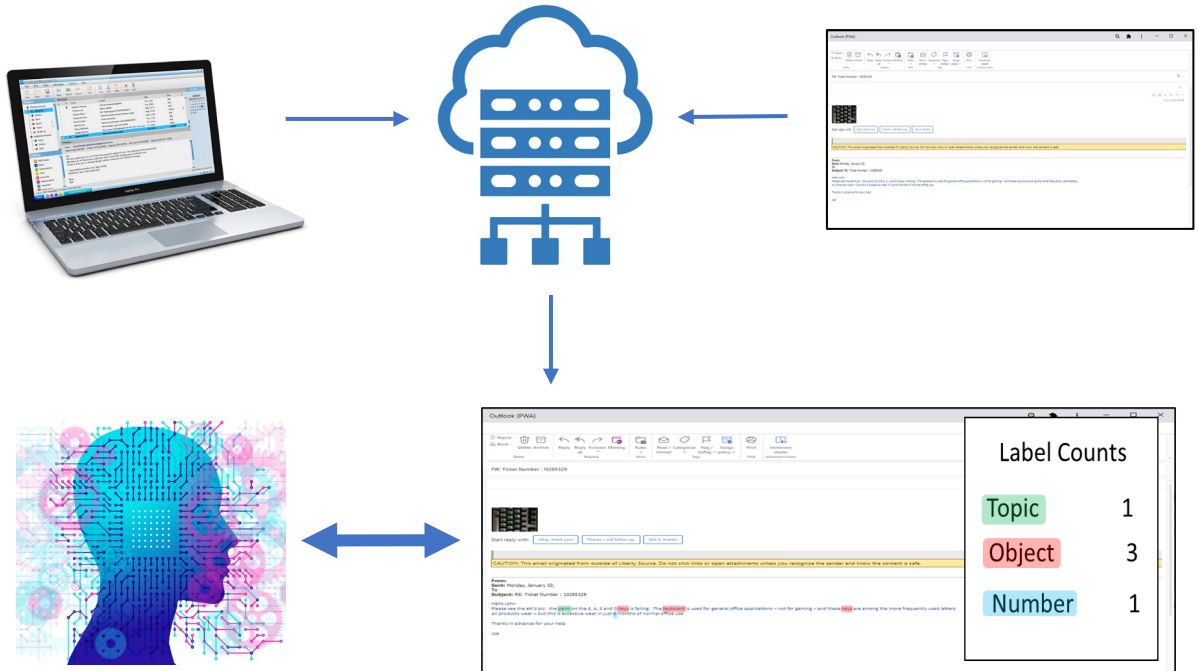
The ability to infer relevance based on context is required in order to successfully harvest all the useful text strings in each document – and that judgment relies on a human-augmented process. Our Data Associates reviewed a wide variety of document formats and file types received by our client’s staff from various sources around the world. While navigating the different layouts and detail levels, Associates parsed out the relevant information from thousands of emails per day, categorized each string, and researched key items to add additional value to the parsed content.

### Outcome

Combining human judgment and context sensitivity with traditional text annotation yielded a customized dataset which allowed our client to realize tremendous insight and significant value from seemingly random external documents. Progress tracking allowed the client to maintain visibility into the volume of documents being processed, while active collaboration ensured that the capture and categorization of data was tightly aligned with the analytical objectives.

## Corporate Email Content Mining

*Source: Inbox messages relating to specific topics*



### **Human in the Loop** –

*Correct labeling of relevant words / strings is verified by Data Associates who utilize judgment and contextual recognition to significantly increase the accuracy of automated labeling platforms.*

```

6 +     selectors: 'text-tag',
7 +     templateUrl: './text-tag.component.html',
8 +     styleUrls: ['./text-tag.component.css']
9 + })
10 + export class TextTagComponent implements OnInit {
11 +   @Input() text: string;
12 +   @Input() tags: string[];
13 +
14 +   document: Document;
15 +   textArray: string[];
16 +   tagData = { 'tagA': [], 'tagB': [], 'tagC': [], 'tagD': [] };
17 +   tagDataArray = [ 'tagA': [], 'tagB': [], 'tagD': [], 'tagC': [] ];
18 +   htmlOfText = [];
19 +   selectedTag: string;
20 +   selectedTagColor: string;
21 +   selection: Selection;
22 +
23 +   ngOnInit() { }
24 +
25 +   ngOnChanges() {
26 +     this.textArray = this.text.split(' ');
27 +   }
    
```

**Output: Code gathers labeled strings into structured dataset for analysis**