**Extracting Json data from AWS Textract.**

**Goal:** Transforming AWS Textract’s csvs into standardized Json files.

AWS Textract is a tool that is similar to Avarni’s crawler. It takes sustainability reports and generates relevant short reports in the csv format.

We have also explored this tool (along with Avarni’s crawler) and developed an application that transforms these csv’s into standardised Json files. This application will be handy if team at Avarni wants to use AWS Textract tool in the future.

We have tested this app with reports from 3 companies (Amazon,Attlassian,Microsoft) that are not entirely similar and tried to produce standardized Json files.

*Challenges and Assumptions:*

1. This will work for all reports that are in one of these three formats.
2. Only scope 1, scope 2 and scope 3 are considered.
3. This is not integrated with AWS Textract ; sustainability reports need to be fed into AWS Textract manually and csv’s outputted by the Textract need to be fed into this application manually.

*Tech Spec of the application:*

Java 11, MVN 3.x,Docker

**Example csv’s generated by the AWS Textract.**

**Attlasian:**

***""***,***"*FY19 *"***,***"*% OF TOTAL *"***,***"*FY20 *"***,***"*% OF TOTAL *"***,  
***"*SCOPE 1 *"***,***"*119.4 *"***,***"*0.14% *"***,***"*274.9 *"***,***"*0.4% *"***,  
***"*Natural Gas *"***,***"*117.8 *"***,***"*0.13% *"***,***"*186.8 *"***,***"*0.2% *"***,  
***"*Diesel *"***,***"*1.6 *"***,***"*0.00% *"***,***"*88.1 *"***,***"*0.1% *"***,  
***"*SCOPE 2 Market-based *"***,***"*2,559 *"***,***"*2.92% *"***,***"*693 *"***,***"*0.9% *"***,  
***"*Fugitive emissions \*not covered by RECs *"***,***"*146 *"***,***"*0.17% *"***,***"*200.9 *"***,***"*0.3% *"***,  
***"*Purchased electricity \*covered by RECs *"***,***"*1,940 *"***,***"*2.21% *"***,***"*2,308.40 *"***,***"*3.0% *"***,  
***"*Purchased heating \*not covered by RECs *"***,***"*394 *"***,***"*0.45% *"***,***"*491.9 *"***,***"*0.6% *"***,  
***"*Colocation datacenters \*covered by RECs *"***,***"*78 *"***,***"*0.09% *"***,***"*78 *"***,***"*0.1% *"***,  
***"*Location-based *"***,***"*2,505 *"***,***"*- *"***,***"*3,099 *"***,***"*- *"***,  
***"*Fugitive emissions *"***,***"*146 *"***,***"*- *"***,***"*200.9 *"***,***"*- *"***,  
***"*Purchased electricity *"***,***"*1,887 *"***,***"*- *"***,***"*2,328.30 *"***,***""***,  
***"*Purchased heating *"***,***"*394 *"***,***"*- *"***,***"*491.9 *"***,***"*- *"***,  
***"*Colocation datacenters *"***,***"*78 *"***,***"*- *"***,***"*78 *"***,***"*- *"***,  
***"*SCOPE 3 *"***,***""***,***""***,***""***,***""***,  
***"*Purchased Goods and Services *"***,***"*44,214 *"***,***"*50.44% *"***,***"*52,376 *"***,***"*69.2% *"***,  
***"*Capital Goods *"***,***"*11,968 *"***,***"*13.65% *"***,***"*7,832 *"***,***"*10.3% *"***,  
***"*Fuel and energy *"***,***"*573 *"***,***"*0.65% *"***,***"*605 *"***,***"*0.8% *"***,  
***"*Transportation *"***,***"*92 *"***,***"*0.1% *"***,***"*74 *"***,***"*0.1% *"***,  
***"*Waste *"***,***"*124 *"***,***"*0.14% *"***,***"*298 *"***,***"*0.4% *"***,  
***"*Business travel *"***,***"*21,702 *"***,***"*24.76% *"***,***"*8,320 *"***,***"*11.0% *"***,  
***"*Commuting *"***,***"*6,308 *"***,***"*7.20% *"***,***"*5,216 *"***,***"*6.9% *"***,  
***"*Leased Assets *"***,***"*31 *"***,***"*0.04% *"***,***"*25 *"***,***"*0.03% *"***,  
***"*Upstream Total *"***,***"*84,982 *"***,***"*96.95% *"***,***"*74,721 *"***,***"*98.7% *"***,  
***"*TOTAL INVENTORY *"***,***"*87,660 *"***,***""***,***"*75,688.90 *"***,***""***,

**Corresponding JSON:**

{

"scope 1\*":{

"FY19 ":{

"Diesel ":1.6,

"Natural Gas ":117.8,

"total":119.4

},

"FY20 ":{

"Diesel ":88.1,

"Natural Gas ":186.8,

"total":274.9

}

},

"scope 2\*":{

"FY19 ":{

"Colocation datacenters ":78.0,

"Colocation datacenters \*covered by RECs ":78.0,

"Fugitive emissions ":146.0,

"Fugitive emissions \*not covered by RECs ":146.0,

"Location-based ":2505.0,

"Purchased electricity ":1887.0,

"Purchased electricity \*covered by RECs ":1940.0,

"Purchased heating ":394.0,

"Purchased heating \*not covered by RECs ":394.0,

"total":7568.0

},

"FY20 ":{

"Colocation datacenters ":78.0,

"Colocation datacenters \*covered by RECs ":78.0,

"Fugitive emissions ":200.9,

"Fugitive emissions \*not covered by RECs ":200.9,

"Location-based ":3099.0,

"Purchased electricity ":2328.3,

"Purchased electricity \*covered by RECs ":2308.4,

"Purchased heating ":491.9,

"Purchased heating \*not covered by RECs ":491.9,

"total":9277.3

}

},

"scope 3\*":{

"FY19 ":{

"Business travel ":21702.0,

"Capital Goods ":11968.0,

"Commuting ":6308.0,

"Fuel and energy ":573.0,

"Leased Assets ":31.0,

"Purchased Goods and Services ":44214.0,

"TOTAL INVENTORY ":87660.0,

"Transportation ":92.0,

"Upstream Total ":84982.0,

"Waste ":124.0,

"total":257654.0

},

"FY20 ":{

"Business travel ":8320.0,

"Capital Goods ":7832.0,

"Commuting ":5216.0,

"Fuel and energy ":605.0,

"Leased Assets ":25.0,

"Purchased Goods and Services ":52376.0,

"TOTAL INVENTORY ":75688.9,

"Transportation ":74.0,

"Upstream Total ":74721.0,

"Waste ":298.0,

"total":225155.9

}

}

}

**Amazon:**

***""***,***"*2018 *"***,***"*2019 *"***,***"*2020 *"***,***"*YOY% *"***,  
***"*Carbon Intensity (grams of CO2e per $ of GMS) *"***,***"*128.9 *"***,***"*122.8 *"***,***"*102.7 *"***,***"*-16% *"***,  
***"*Emissions Category (mmt CO2e) *"***,***"*2018 *"***,***"*2019 *"***,***"*2020 *"***,***"*YOY% *"***,  
***"*Emissions from Direct Operations (Scope 1) *"***,***"*4.98 *"***,***"*5.76 *"***,***"*9.62 *"***,***"*+67% *"***,  
***"*Fossil fuels *"***,***"*4.70 *"***,***"*5.57 *"***,***"*9.37 *"***,***"*+69% *"***,  
***"*Refrigerants *"***,***"*0.28 *"***,***"*0.19 *"***,***"*0.25 *"***,***"*+32% *"***,  
***"*Emissions from Purchased Electricity (Scope 2) *"***,***"*4.71 *"***,***"*5.50 *"***,***"*5.27 *"***,***"*-4% *"***,  
***"*Emissions from Indirect Sources (Scope 3) *"***,***"*34.71 *"***,***"*39.91 *"***,***"*45.75 *"***,***"*+15% *"***,  
***"*Corporate purchases and Amazon-branded product emissions (e.g., operating expenses, business travel, and Amazon-branded product manufacturing, use phase, and end-of-life) *"***,***"*11.95 *"***,***"*15.41 *"***,***"*16.70 *"***,***"*+8% *"***,  
***"*Capital Goods (e.g., building construction, servers and other hardware, equipment, vehicles) *"***,***"*4.64 *"***,***"*8.01 *"***,***"*10.52 *"***,***"*+31% *"***,  
***"*Other indirect emissions (e.g., third-party transportation, packaging, upstream energy related) *"***,***"*13.89 *"***,***"*12.44 *"***,***"*15.77 *"***,***"*+27% *"***,  
***"*Lifecycle emissions from customer trips to Amazon's physical stores *"***,***"*4.23 *"***,***"*4.05 *"***,***"*2.77 *"***,***"*-32% *"***,  
***"*Amazon's Total Footprint *"***,***"*44.40 *"***,***"*51.17 *"***,***"*60.64 *"***,***"*+19% *"***,

**Corresponding JSON:**

{

"scope 1\*":{

"2018 ":{

"Emissions from Direct Operations (Scope 1) ":4.98,

"Fossil fuels ":4.7,

"Refrigerants ":0.28,

"total":9.96

},

"2019 ":{

"Emissions from Direct Operations (Scope 1) ":5.76,

"Fossil fuels ":5.57,

"Refrigerants ":0.19,

"total":11.52

},

"2020 ":{

"Emissions from Direct Operations (Scope 1) ":9.62,

"Fossil fuels ":9.37,

"Refrigerants ":0.25,

"total":19.24

}

},

"scope 2\*":{

"2018 ":{

"Emissions from Purchased Electricity (Scope 2) ":4.71,

"total":4.71

},

"2019 ":{

"Emissions from Purchased Electricity (Scope 2) ":5.5,

"total":5.5

},

"2020 ":{

"Emissions from Purchased Electricity (Scope 2) ":5.27,

"total":5.27

}

},

"scope 3\*":{

"2018 ":{

"Amazon's Total Footprint ":44.4,

"Capital Goods (e.g., building construction, servers and other hardware, equipment, vehicles) ":4.64,

"Corporate purchases and Amazon-branded product emissions (e.g., operating expenses, business travel, and Amazon-branded product manufacturing, use phase, and end-of-life) ":11.95,

"Emissions from Indirect Sources (Scope 3) ":34.71,

"Lifecycle emissions from customer trips to Amazon's physical stores ":4.23,

"Other indirect emissions (e.g., third-party transportation, packaging, upstream energy related) ":13.89,

"total":113.82

},

"2019 ":{

"Amazon's Total Footprint ":51.17,

"Capital Goods (e.g., building construction, servers and other hardware, equipment, vehicles) ":8.01,

"Corporate purchases and Amazon-branded product emissions (e.g., operating expenses, business travel, and Amazon-branded product manufacturing, use phase, and end-of-life) ":15.41,

"Emissions from Indirect Sources (Scope 3) ":39.91,

"Lifecycle emissions from customer trips to Amazon's physical stores ":4.05,

"Other indirect emissions (e.g., third-party transportation, packaging, upstream energy related) ":12.44,

"total":130.99

},

"2020 ":{

"Amazon's Total Footprint ":60.64,

"Capital Goods (e.g., building construction, servers and other hardware, equipment, vehicles) ":10.52,

"Corporate purchases and Amazon-branded product emissions (e.g., operating expenses, business travel, and Amazon-branded product manufacturing, use phase, and end-of-life) ":16.7,

"Emissions from Indirect Sources (Scope 3) ":45.75,

"Lifecycle emissions from customer trips to Amazon's physical stores ":2.77,

"Other indirect emissions (e.g., third-party transportation, packaging, upstream energy related) ":15.77,

"total":152.15

}

}

}

**Microsoft**:

***""***,***"*FY17 *"***,***"*FY18 *"***,***"*FY19 *"***,***"*FY20 *"***,  
***"*Scope 11,2 *"***,***"*107,452 *"***,***"*99,008 *"***,***"*117,956 *"***,***"*118,100 *"***,  
***""***,***""***,***""***,***""***,***""***,  
***"*Scope 1 Scope1-based² *"***,***"*2,697,554 *"***,***"*10000 *"***,***"*3,557,518 *"***,***"*4,102,445 *"***,  
***"*Scope 1 Scope1-based1,2 *"***,***"*139,066 *"***,***"*11000 *"***,***"*275,420 *"***,***"*228,194 *"***,  
***"*Scope 2 Location-based² *"***,***"*2,697,554 *"***,***"*2,946,043 *"***,***"*3,557,518 *"***,***"*4,102,445 *"***,  
***"*Scope 2 Market-based1,2 *"***,***"*139,066 *"***,***"*183,329 *"***,***"*275,420 *"***,***"*228,194 *"***,  
***"*Scope 3 - Purchased goods & services3,5 *"***,***"*4,090,000 *"***,***"*4,120,000 *"***,***"*4,190,000 *"***,***"*4,050,000 *"***,  
***"*Scope 3 Capital goods3,5 *"***,***"*1,670,000 *"***,***"*2,230,000 *"***,***"*2,180,000 *"***,***"*2,750,000 *"***,  
***"*Scope 3 - Fuel- and energy-related activities (Location-based)5 *"***,***"*540,000 *"***,***"*550,000 *"***,***"*650,000 *"***,***"*770,000 *"***,  
***"*Scope 3 - Fuel- and energy-related activities (Market-based)5 *"***,***"*89,000 *"***,***"*110,000 *"***,***"*170,000 *"***,***"*180,000 *"***,  
***"*Scope 3 - Upstream transportation³, *"***,***"*50,000 *"***,***"*50,000 *"***,***"*100,000 *"***,***"*100,000 *"***,  
***"*Scope 3 Waste5 *"***,***"*700 *"***,***"*500 *"***,***"*1,100 *"***,***"*800 *"***,  
***"*Scope 3 Business travel1,4 *"***,***"*419,020 *"***,***"*461,787 *"***,***"*476,457 *"***,***"*329,356 *"***,  
***"*Scope 3 - Employee commuting2,5 *"***,***"*343,000 *"***,***"*345,000 *"***,***"*411,000 *"***,***"*317,000 *"***,  
***"*Scope 3 - Downstream transportation2,5 *"***,***"*85,000 *"***,***"*98,000 *"***,***"*57,000 *"***,***"*44,000 *"***,  
***"*Scope 3 - Use of sold products2,5 *"***,***"*4,103,000 *"***,***"*4,230,000 *"***,***"*3,753,000 *"***,***"*3,025,000 *"***,  
***"*Scope 3 - End-of-life of sold products2,5 *"***,***"*31,000 *"***,***"*18,000 *"***,***"*18,000 *"***,***"*17,000 *"***,  
***"*Scope 3 - Downstream leased assets *"***,***"*700 *"***,***"*1,700 *"***,***"*800 *"***,***"*4,600 *"***,  
***"*Total Scope 1, Scope 2, and Scope 36 *"***,***"*11,128,000 *"***,***"*11,947,000 *"***,***"*11,751,000 *"***,***"*11,164,000 *"***,

**Corresponding JSON:**

{

"scope 1\*":{

"FY17 ":{

"Scope 1 Scope1-based1,2 ":139066.0,

"Scope 1 Scope1-based² ":2697554.0,

"Scope 11,2 ":107452.0,

"total":2944072.0

},

"FY18 ":{

"Scope 1 Scope1-based1,2 ":11000.0,

"Scope 1 Scope1-based² ":10000.0,

"Scope 11,2 ":99008.0,

"total":120008.0

},

"FY19 ":{

"Scope 1 Scope1-based1,2 ":275420.0,

"Scope 1 Scope1-based² ":3557518.0,

"Scope 11,2 ":117956.0,

"total":3950894.0

},

"FY20 ":{

"Scope 1 Scope1-based1,2 ":228194.0,

"Scope 1 Scope1-based² ":4102445.0,

"Scope 11,2 ":118100.0,

"total":4448739.0

}

},

"scope 2\*":{

"FY17 ":{

"Scope 2 Location-based² ":2697554.0,

"Scope 2 Market-based1,2 ":139066.0,

"total":2836620.0

},

"FY18 ":{

"Scope 2 Location-based² ":2946043.0,

"Scope 2 Market-based1,2 ":183329.0,

"total":3129372.0

},

"FY19 ":{

"Scope 2 Location-based² ":3557518.0,

"Scope 2 Market-based1,2 ":275420.0,

"total":3832938.0

},

"FY20 ":{

"Scope 2 Location-based² ":4102445.0,

"Scope 2 Market-based1,2 ":228194.0,

"total":4330639.0

}

},

"scope 3\*":{

"FY17 ":{

"Scope 3 - Downstream leased assets ":700.0,

"Scope 3 - Downstream transportation2,5 ":85000.0,

"Scope 3 - Employee commuting2,5 ":343000.0,

"Scope 3 - End-of-life of sold products2,5 ":31000.0,

"Scope 3 - Fuel- and energy-related activities (Location-based)5 ":540000.0,

"Scope 3 - Fuel- and energy-related activities (Market-based)5 ":89000.0,

"Scope 3 - Purchased goods & services3,5 ":4090000.0,

"Scope 3 - Upstream transportation³, ":50000.0,

"Scope 3 - Use of sold products2,5 ":4103000.0,

"Scope 3 Business travel1,4 ":419020.0,

"Scope 3 Capital goods3,5 ":1670000.0,

"Scope 3 Waste5 ":700.0,

"total":1.142142E7

},

"FY18 ":{

"Scope 3 - Downstream leased assets ":1700.0,

"Scope 3 - Downstream transportation2,5 ":98000.0,

"Scope 3 - Employee commuting2,5 ":345000.0,

"Scope 3 - End-of-life of sold products2,5 ":18000.0,

"Scope 3 - Fuel- and energy-related activities (Location-based)5 ":550000.0,

"Scope 3 - Fuel- and energy-related activities (Market-based)5 ":110000.0,

"Scope 3 - Purchased goods & services3,5 ":4120000.0,

"Scope 3 - Upstream transportation³, ":50000.0,

"Scope 3 - Use of sold products2,5 ":4230000.0,

"Scope 3 Business travel1,4 ":461787.0,

"Scope 3 Capital goods3,5 ":2230000.0,

"Scope 3 Waste5 ":500.0,

"total":1.2214987E7

},

"FY19 ":{

"Scope 3 - Downstream leased assets ":800.0,

"Scope 3 - Downstream transportation2,5 ":57000.0,

"Scope 3 - Employee commuting2,5 ":411000.0,

"Scope 3 - End-of-life of sold products2,5 ":18000.0,

"Scope 3 - Fuel- and energy-related activities (Location-based)5 ":650000.0,

"Scope 3 - Fuel- and energy-related activities (Market-based)5 ":170000.0,

"Scope 3 - Purchased goods & services3,5 ":4190000.0,

"Scope 3 - Upstream transportation³, ":100000.0,

"Scope 3 - Use of sold products2,5 ":3753000.0,

"Scope 3 Business travel1,4 ":476457.0,

"Scope 3 Capital goods3,5 ":2180000.0,

"Scope 3 Waste5 ":1100.0,

"total":1.2007357E7

},

"FY20 ":{

"Scope 3 - Downstream leased assets ":4600.0,

"Scope 3 - Downstream transportation2,5 ":44000.0,

"Scope 3 - Employee commuting2,5 ":317000.0,

"Scope 3 - End-of-life of sold products2,5 ":17000.0,

"Scope 3 - Fuel- and energy-related activities (Location-based)5 ":770000.0,

"Scope 3 - Fuel- and energy-related activities (Market-based)5 ":180000.0,

"Scope 3 - Purchased goods & services3,5 ":4050000.0,

"Scope 3 - Upstream transportation³, ":100000.0,

"Scope 3 - Use of sold products2,5 ":3025000.0,

"Scope 3 Business travel1,4 ":329356.0,

"Scope 3 Capital goods3,5 ":2750000.0,

"Scope 3 Waste5 ":800.0,

"total":1.1587756E7

}

}

}