

INFO 300 Assignment 2

1. Student(s)

- Weimao Ke, wk77@drexel.edu
- John Smith, jps2020@drexel.edu

2. Tasks and Steps

2.1 Text Collection (2 points)

Example article:

```
{
  "author": "Williams, F (Williams, Franco)",
  "title": "Decarbonisation pathways of the cement production process via
hydrogen and oxy-combustion",
  "abstract": "Decarbonising cement production is of profound importance for
meeting global greenhouse gas emission reduction targets and mitigating the impact
of climate change. This study evaluates various technical options for achieving
deep decarbonisation in a clinker production facility by utilising hydrogen (H2)
as an alternative fuel to replace fossil fuels and by integrating an oxy-
combustion technique with carbon capture and storage (CCS). Using Aspen Plus
process simulations, we examined the extent of decarbonisation and assessed the
thermal and electrical energy demands. This was achieved by incorporating an
amine-absorption-based CO2 capture to a conventional natural gas fuelled reference
plant, implementing oxyfuel-combustion of natural gas, and exploring four
different scenarios for replacing fossil fuel with H2. In these scenarios, H2 was
assumed to be produced through on-site water electrolysis, which also supplied
oxygen for oxyfuel combustion, potentially eliminating the need for an air
separation unit (ASU). The processes utilizing H2, except for the case of
indirectly heated precalcination, employed oxyfuel combustion. The results
indicate that the natural gas-fuelled oxyfuel-combustion process had the lowest
total energy input at 4.92 GJ/t clinker, approximately 35% lower than that of the
reference plant. Processes using H2 reduced energy demand by 11% in the H2-d
scenario and 33% in the H2-a scenario. However, the process with indirect
calcination required 6.24 GJ/t clinker, about 8% more H2 fuel than direct
calcination but helped eliminate the need for an ASU. The results also reveal that
greater H2 substitutions led to higher total process energy requirements due to
the inefficiencies of the electrolysis process. While the H2using processes could
reduce the CO2 generation by up to 559 kgCO2/t clinker, this represents only about
27.6% of the CO2 reductions relative to the reference plant. These findings
underscore the limitation of fuel substitution alone in cement production and
emphasize the need for innovations in raw materials and the adoption of CCS to
achieve deeper decarbonisation in cement industries.",
  "citations": 6
}
```

2.2. Index Mapping (2 points)

Mapping Request 1

Request 1:

```
PUT /zhangqirui_articles
{
  "mappings": {
    "properties": {
      "author": {
        "type": "text"
      },
      "title": {
        "type": "text",
        "analyzer": "english"
      },
      "abstract": {
        "type": "text",
        "analyzer": "english"
      },
      "citations": {
        "type": "integer"
      }
    }
  }
}
```

Response 1:

```
{
  "acknowledged": true,
  "shards_acknowledged": true,
  "index": "zhangqirui_articles"
}
```

Mapping Request 2

Request:

```
PUT /zhangqirui_articles/_mapping
{
  "properties": {
    "author": {
      "type": "text"
    },
    "title": {
      "type": "text",
      "analyzer": "english"
    },
  },
}
```

```
"abstract": {
  "type": "text",
  "analyzer": "english"
},
"citations": {
  "type": "integer"
}
}
```

Response:

```
{
  "acknowledged": true
}
```

2.3. Data Indexing (1 point)

Request:

```
PUT /zhangqirui_articles/_doc/1
{
  "author": "Williams, F (Williams, Franco)",
  "title": "Decarbonisation pathways of the cement production process via hydrogen and oxy-combustion",
  "abstract": "Decarbonising cement production is of profound importance for meeting global greenhouse gas emission reduction targets and mitigating the impact of climate change. This study evaluates various technical options for achieving deep decarbonisation in a clinker production facility by utilising hydrogen (H2) as an alternative fuel to replace fossil fuels and by integrating an oxy-combustion technique with carbon capture and storage (CCS). Using Aspen Plus process simulations, we examined the extent of decarbonisation and assessed the thermal and electrical energy demands. This was achieved by incorporating an amine-absorption-based CO2 capture to a conventional natural gas fuelled reference plant, implementing oxyfuel-combustion of natural gas, and exploring four different scenarios for replacing fossil fuel with H2. In these scenarios, H2 was assumed to be produced through on-site water electrolysis, which also supplied oxygen for oxyfuel combustion, potentially eliminating the need for an air separation unit (ASU). The processes utilizing H2, except for the case of indirectly heated precalcination, employed oxyfuel combustion. The results indicate that the natural gas-fuelled oxyfuel-combustion process had the lowest total energy input at 4.92 GJ/t clinker, approximately 35% lower than that of the reference plant. Processes using H2 reduced energy demand by 11% in the H2-d scenario and 33% in the H2-a scenario. However, the process with indirect calcination required 6.24 GJ/t clinker, about 8% more H2 fuel than direct calcination but helped eliminate the need for an ASU. The results also reveal that greater H2 substitutions led to higher total process energy requirements due to the inefficiencies of the electrolysis process. While the H2using processes could reduce the CO2 generation by up to 559 kgCO2/t clinker, this represents only about
```

```
27.6% of the CO2 reductions relative to the reference plant. These findings
underscore the limitation of fuel substitution alone in cement production and
emphasize the need for innovations in raw materials and the adoption of CCS to
achieve deeper decarbonisation in cement industries.",
  "citations": 6
}
```

Response:

```
{
  "_index": "zhangqirui_articles",
  "_id": "1",
  "_version": 1,
  "result": "created",
  "_shards": {
    "total": 2,
    "successful": 1,
    "failed": 0
  },
  "_seq_no": 0,
  "_primary_term": 1
}
```

2.4. Search and Retrieval (1 point)

Request:

```
GET /zhangqirui_articles/_search
{
  "query": {
    "match": {
      "abstract": "renewable power supply"
    }
  }
}
```

Response (**top three** articles):

```
{
  "_index": "zhangqirui_articles",
  "_id": "15",
  "_score": 5.9451346,
  "_source": {
    "author": "Xing,Chen, C (Chen, Chao)",
    "title": "Renewable methanol production: Understanding the interplay
between storage sizing, renewable mix and dispatchable energy price",
    "abstract": "Chemical production using renewable energies is an
```

important element on the roadmap of industry decarbonisation. This work investigates the optimisation of renewable power supply for a fully electrified methanol process, with a focus on the interplay between renewable fix, storage sizing and the use of backup dispatchable power source. The analysis is performed using the meteorological data obtained from two locations, i.e. Kramer Junction (US) and Norderney (Germany), which have excellent solar and wind source, respectively. The minimum levelised energy cost, which is optimised in terms of renewable power generation, renewable mix and storage size, is found to be 106\$/MWh and 103\$/MWh for operations in Kramer Junction and Norderney, respectively, based on a dispatchable energy price of 230\$/MWh. This leads to a levelised methanol cost of 1490\$/tonne and 1459\$/tonne with a respective renewable penetration of 81% and 96% in the production. The correlation between renewable penetration and dispatchable energy price in the most economical scenario exhibits a two-regime behaviour: the renewable penetration increases dramatically at the beginning and then slowly approaches 100% when the dispatchable energy price is above a critical point. For a fully renewable operation, the optimised levelised energy cost is found to increase to 167\$/MWh and 114\$/MWh for Kramer Junction and Norderney, respectively. The results show the importance of the dual functionality of hydrogen in the energy storage system, which improves the overall energy efficiency."

"citations": 31

}

},

{

"_index": "zhangqirui_articles",

"_id": "18",

"_score": 5.3530884,

"_source": {

"author": "Chen, C (Chen, Chao)",

"title": "Power-to-methanol: The role of process flexibility in the integration of variable renewable energy into chemical production",

"abstract": "Chemical process electrification and renewable energy integration facilitate one another along the pathway towards a greener industry. However, integrating intermittent and variable renewable power into large-scale chemical processes, which conventionally are preferred to operate at a steady-state with a constant load, could lead to prohibitive costs if intermittency is addressed solely by energy storage. Here, we consider the concept of a flexible chemical process which can operate with a variable load throughout the year while meeting a specified annual production target. Using methanol production via carbon dioxide hydrogenation as a case study and by means of process conceptual design and optimisation, we investigate how the over-sizing of flexible process units and the introduction of intermediate storage in the chemical process offer the possibility to improve the overall performance of systems. The impact of the characteristics of renewable power is also explored by performing the analysis using meteorological data from two locations dominated respectively by wind and solar energy. This study shows clear potential benefits of process flexibility when the renewable energy supply is highly variable and is to achieve a high level of penetration. For a 100% renewable production, the introduction of flexibility reduces the levelised cost of methanol by approximately 21 and 34% for the two case study locations, respectively. The cost attribution reveals further insights into the origin of the economic advantages through examining the comparative costs of chemical production, energy generation, intermediate product storage and renewable energy storage. The learning from this work suggests that incorporating process flexibility through a holistically optimised design of energy storage and

```

chemical production has the potential to offer an economically viable route to
large-scale green chemical production through renewables-enabled
electrification.",
  "citations": 73
},
{
  "_index": "zhangqirui_articles",
  "_id": "7",
  "_score": 4.464466,
  "_source": {
    "author": "Devlin, A (Devlin, Alexandra)",
    "title": "Regional supply chains for decarbonising steel: Energy
efficiency and green premium mitigation",
    "abstract": "Decarbonised steel, enabled by green hydrogen-based iron
ore reduction and renewable electricity-based steel making, will disrupt the
traditional supply chain. Focusing on the energetic and techno-economic assessment
of potential green supply chains, this study investigates the direct reduced iron-
electric arc furnace production route enabled by renewable energy and deployed in
regional settings. The hypothesis, that co-locating manufacturing processes with
renewable energy resources would offer highest energy efficiency and cost
reduction, is tested through an Australia-Japan case study. The binational
partnership is structured to meet Japanese steel demand (for domestic use and
regional exports) and source both energy and iron ore from the Pilbara region of
Western Australia. A total of 12 unique supply chains differentiated by spatial
configuration, timeline and energy carrier were simulated, which validated the
hypothesis: direct energy and ore exports to remote steel producers (i.e. Japan-
based production), as opposed to co-locating iron and steel production with
abundant ore and renewable energy resources (i.e. Australia-based production),
increased energy consumption and the levelised cost of steel by 45% and 32%,
respectively, when averaged across 2030 and 2050. Two decades of technological
development and economies of scale realisation would be crucial; 2030 supply
chains were on average 12% more energy-intense and 23% more expensive than 2050
equivalents. On energy vectors, liquefied hydrogen was more efficient than ammonia
for export-dominant supply chains due to the pairing of its process flexibility
and the intermittent solar energy profile, as well as the avoidance of the need
for ammonia cracking prior to direct reduction. To mitigate the green premium, a
carbon tax in the range of A$66-192/t CO2 would be required in 2030 and A$0-70/t
CO2 in 2050; the diminished carbon tax requirement in the latter is achievable
only by wholly Australia-based production. Further, the modelled system scale was
immense; producing 40 Mtpa of decarbonised steel will require 74-129% of
Australia's current electricity output and A$137-328 billion in capital investment
for solar power, production, and shipping vessel infrastructure. These results
call for strategic planning of regional resource pairing to drive energy and cost
efficiencies which accelerate the global decar-bonisation of steel.",
    "citations": 31
  }
}

```

2.5. Manual Examination

A. Term stats (2 points)

Request for DF Stats:

```
GET /zhangqirui_articles/_termvectors
{
  "doc": {
    "abstract": "renewable power supply"
  },
  "term_statistics": true,
  "field_statistics": false,
  "positions": false,
  "offsets": false
}
```

Response:

```
{
  "_index": "zhangqirui_articles",
  "_version": 0,
  "found": true,
  "took": 0,
  "term_vectors": {
    "abstract": {
      "terms": {
        "power": {
          "doc_freq": 4,
          "ttf": 7,
          "term_freq": 1
        },
        "renew": {
          "doc_freq": 6,
          "ttf": 24,
          "term_freq": 1
        },
        "suppli": {
          "doc_freq": 13,
          "ttf": 23,
          "term_freq": 1
        }
      }
    }
  }
}
```

Request for TF stats in document/hit 1:

```
GET /zhangqirui_articles/_termvectors/15
{
  "fields": ["abstract"],
```

```
"term_statistics": true,  
"field_statistics": false,  
"positions": false,  
"offsets": false  
}
```

Response:

```
{  
  "_index": "zhangqirui_articles",  
  "_id": "15",  
  "_version": 2,  
  "found": true,  
  "took": 2,  
  "term_vectors": {  
    "abstract": {  
      "terms": {  
        "81": {  
          "doc_freq": 1,  
          "ttf": 1,  
          "term_freq": 1  
        },  
        "96": {  
          "doc_freq": 1,  
          "ttf": 1,  
          "term_freq": 1  
        },  
        "100": {  
          "doc_freq": 3,  
          "ttf": 3,  
          "term_freq": 1  
        },  
        "103": {  
          "doc_freq": 1,  
          "ttf": 1,  
          "term_freq": 1  
        },  
        "106": {  
          "doc_freq": 1,  
          "ttf": 1,  
          "term_freq": 1  
        },  
        "114": {  
          "doc_freq": 1,  
          "ttf": 1,  
          "term_freq": 1  
        },  
        "167": {  
          "doc_freq": 1,  
          "ttf": 1,  
          "term_freq": 1  
        },  
      }  
    }  
  }  
}
```



```
"230": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"1459": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"1490": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"abov": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"analysi": {
  "doc_freq": 7,
  "ttf": 9,
  "term_freq": 1
},
"approach": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"backup": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"base": {
  "doc_freq": 14,
  "ttf": 29,
  "term_freq": 1
},
"begin": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"behaviour": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"between": {
  "doc_freq": 11,
  "ttf": 18,
  "term_freq": 2
}
```

```
},
"chemic": {
  "doc_freq": 8,
  "ttf": 17,
  "term_freq": 1
},
"correl": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"cost": {
  "doc_freq": 9,
  "ttf": 26,
  "term_freq": 3
},
"critic": {
  "doc_freq": 2,
  "ttf": 3,
  "term_freq": 1
},
"data": {
  "doc_freq": 7,
  "ttf": 15,
  "term_freq": 1
},
"decarbonis": {
  "doc_freq": 5,
  "ttf": 9,
  "term_freq": 1
},
"dispatch": {
  "doc_freq": 1,
  "ttf": 4,
  "term_freq": 4
},
"dramat": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"dual": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"econom": {
  "doc_freq": 4,
  "ttf": 5,
  "term_freq": 1
},
"effici": {
  "doc_freq": 6,
  "ttf": 10,
```

```
    "term_freq": 1
  },
  "electrifi": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "element": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "energi": {
    "doc_freq": 12,
    "ttf": 64,
    "term_freq": 8
  },
  "excel": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "exhibit": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "fix": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "focu": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "found": {
    "doc_freq": 4,
    "ttf": 5,
    "term_freq": 2
  },
  "from": {
    "doc_freq": 15,
    "ttf": 30,
    "term_freq": 1
  },
  "fulli": {
    "doc_freq": 2,
    "ttf": 3,
    "term_freq": 2
  },
  "function": {
    "doc_freq": 1,
```

```
    "ttf": 1,
    "term_freq": 1
  },
  "gener": {
    "doc_freq": 4,
    "ttf": 4,
    "term_freq": 1
  },
  "germani": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "have": {
    "doc_freq": 10,
    "ttf": 15,
    "term_freq": 1
  },
  "hydrogen": {
    "doc_freq": 5,
    "ttf": 6,
    "term_freq": 1
  },
  "i.": {
    "doc_freq": 6,
    "ttf": 7,
    "term_freq": 1
  },
  "import": {
    "doc_freq": 4,
    "ttf": 8,
    "term_freq": 2
  },
  "improv": {
    "doc_freq": 9,
    "ttf": 11,
    "term_freq": 1
  },
  "increas": {
    "doc_freq": 8,
    "ttf": 10,
    "term_freq": 2
  },
  "industri": {
    "doc_freq": 6,
    "ttf": 6,
    "term_freq": 1
  },
  "interplai": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "investig": {
```

```
    "doc_freq": 6,
    "ttf": 6,
    "term_freq": 1
  },
  "junction": {
    "doc_freq": 1,
    "ttf": 3,
    "term_freq": 3
  },
  "kramer": {
    "doc_freq": 1,
    "ttf": 3,
    "term_freq": 3
  },
  "lead": {
    "doc_freq": 2,
    "ttf": 2,
    "term_freq": 1
  },
  "levelis": {
    "doc_freq": 3,
    "ttf": 5,
    "term_freq": 3
  },
  "locat": {
    "doc_freq": 6,
    "ttf": 15,
    "term_freq": 1
  },
  "meteorolog": {
    "doc_freq": 2,
    "ttf": 2,
    "term_freq": 1
  },
  "methanol": {
    "doc_freq": 2,
    "ttf": 4,
    "term_freq": 2
  },
  "minimum": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "mix": {
    "doc_freq": 2,
    "ttf": 2,
    "term_freq": 1
  },
  "most": {
    "doc_freq": 6,
    "ttf": 7,
    "term_freq": 1
  },
  },
```

```
"mwh": {
  "doc_freq": 1,
  "ttf": 5,
  "term_freq": 5
},
"nordernei": {
  "doc_freq": 1,
  "ttf": 3,
  "term_freq": 3
},
"obtain": {
  "doc_freq": 3,
  "ttf": 3,
  "term_freq": 1
},
"oper": {
  "doc_freq": 7,
  "ttf": 9,
  "term_freq": 2
},
"optimis": {
  "doc_freq": 5,
  "ttf": 8,
  "term_freq": 3
},
"overall": {
  "doc_freq": 5,
  "ttf": 6,
  "term_freq": 1
},
"penetr": {
  "doc_freq": 2,
  "ttf": 4,
  "term_freq": 3
},
"perform": {
  "doc_freq": 8,
  "ttf": 13,
  "term_freq": 1
},
"point": {
  "doc_freq": 2,
  "ttf": 3,
  "term_freq": 1
},
"power": {
  "doc_freq": 4,
  "ttf": 7,
  "term_freq": 3
},
"price": {
  "doc_freq": 2,
  "ttf": 4,
  "term_freq": 3
}
```

```
},
"process": {
  "doc_freq": 6,
  "ttf": 21,
  "term_freq": 1
},
"product": {
  "doc_freq": 15,
  "ttf": 44,
  "term_freq": 2
},
"regim": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"renew": {
  "doc_freq": 6,
  "ttf": 24,
  "term_freq": 9
},
"respect": {
  "doc_freq": 4,
  "ttf": 8,
  "term_freq": 4
},
"result": {
  "doc_freq": 15,
  "ttf": 23,
  "term_freq": 1
},
"roadmap": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"scenario": {
  "doc_freq": 2,
  "ttf": 5,
  "term_freq": 1
},
"show": {
  "doc_freq": 12,
  "ttf": 13,
  "term_freq": 1
},
"size": {
  "doc_freq": 3,
  "ttf": 4,
  "term_freq": 2
},
"slowli": {
  "doc_freq": 1,
  "ttf": 1,
```

```
    "term_freq": 1
  },
  "solar": {
    "doc_freq": 5,
    "ttf": 6,
    "term_freq": 1
  },
  "sourc": {
    "doc_freq": 3,
    "ttf": 4,
    "term_freq": 2
  },
  "storag": {
    "doc_freq": 3,
    "ttf": 9,
    "term_freq": 3
  },
  "suppli": {
    "doc_freq": 13,
    "ttf": 23,
    "term_freq": 1
  },
  "system": {
    "doc_freq": 13,
    "ttf": 25,
    "term_freq": 1
  },
  "term": {
    "doc_freq": 3,
    "ttf": 3,
    "term_freq": 1
  },
  "tonn": {
    "doc_freq": 2,
    "ttf": 4,
    "term_freq": 2
  },
  "two": {
    "doc_freq": 5,
    "ttf": 8,
    "term_freq": 2
  },
  "us": {
    "doc_freq": 16,
    "ttf": 32,
    "term_freq": 4
  },
  "when": {
    "doc_freq": 4,
    "ttf": 4,
    "term_freq": 1
  },
  "which": {
    "doc_freq": 16,
```



```

      "ttf": 28,
      "term_freq": 3
    },
    "wind": {
      "doc_freq": 4,
      "ttf": 4,
      "term_freq": 1
    },
    "work": {
      "doc_freq": 8,
      "ttf": 8,
      "term_freq": 1
    }
  }
}
}
}

```

Request for TF stats in document/hit 2:

```

GET /zhangqirui_articles/_termvectors/18
{
  "fields": ["abstract"],
  "term_statistics": true,
  "field_statistics": false,
  "positions": false,
  "offsets": false
}

```

Response:

```

{
  "_index": "zhangqirui_articles",
  "_id": "18",
  "_version": 1,
  "found": true,
  "took": 1,
  "term_vectors": {
    "abstract": {
      "terms": {
        "21": {
          "doc_freq": 1,
          "ttf": 1,
          "term_freq": 1
        },
        "34": {
          "doc_freq": 1,
          "ttf": 1,
          "term_freq": 1
        }
      }
    }
  }
}

```

```
},
"100": {
  "doc_freq": 3,
  "ttf": 3,
  "term_freq": 1
},
"achiev": {
  "doc_freq": 6,
  "ttf": 8,
  "term_freq": 1
},
"address": {
  "doc_freq": 5,
  "ttf": 5,
  "term_freq": 1
},
"advantag": {
  "doc_freq": 2,
  "ttf": 2,
  "term_freq": 1
},
"along": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"also": {
  "doc_freq": 7,
  "ttf": 8,
  "term_freq": 1
},
"analysi": {
  "doc_freq": 7,
  "ttf": 9,
  "term_freq": 1
},
"annual": {
  "doc_freq": 3,
  "ttf": 3,
  "term_freq": 1
},
"anoth": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"approxim": {
  "doc_freq": 3,
  "ttf": 3,
  "term_freq": 1
},
"attribut": {
  "doc_freq": 2,
  "ttf": 2,
```

```
    "term_freq": 1
  },
  "benefit": {
    "doc_freq": 2,
    "ttf": 2,
    "term_freq": 1
  },
  "can": {
    "doc_freq": 8,
    "ttf": 11,
    "term_freq": 1
  },
  "carbon": {
    "doc_freq": 6,
    "ttf": 8,
    "term_freq": 1
  },
  "case": {
    "doc_freq": 5,
    "ttf": 6,
    "term_freq": 2
  },
  "characterist": {
    "doc_freq": 2,
    "ttf": 2,
    "term_freq": 1
  },
  "chemic": {
    "doc_freq": 8,
    "ttf": 17,
    "term_freq": 7
  },
  "clear": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "compar": {
    "doc_freq": 6,
    "ttf": 9,
    "term_freq": 1
  },
  "concept": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "conceptu": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "consid": {
    "doc_freq": 6,
```

```
    "ttf": 7,
    "term_freq": 1
  },
  "constant": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "convention": {
    "doc_freq": 4,
    "ttf": 4,
    "term_freq": 1
  },
  "cost": {
    "doc_freq": 9,
    "ttf": 26,
    "term_freq": 4
  },
  "could": {
    "doc_freq": 11,
    "ttf": 18,
    "term_freq": 1
  },
  "data": {
    "doc_freq": 7,
    "ttf": 15,
    "term_freq": 1
  },
  "design": {
    "doc_freq": 4,
    "ttf": 5,
    "term_freq": 2
  },
  "dioxid": {
    "doc_freq": 2,
    "ttf": 3,
    "term_freq": 1
  },
  "domin": {
    "doc_freq": 3,
    "ttf": 3,
    "term_freq": 1
  },
  "econom": {
    "doc_freq": 4,
    "ttf": 5,
    "term_freq": 2
  },
  "electrif": {
    "doc_freq": 1,
    "ttf": 2,
    "term_freq": 2
  },
  "enabl": {
```

```
    "doc_freq": 3,
    "ttf": 4,
    "term_freq": 1
  },
  "energi": {
    "doc_freq": 12,
    "ttf": 64,
    "term_freq": 7
  },
  "examin": {
    "doc_freq": 4,
    "ttf": 4,
    "term_freq": 1
  },
  "explor": {
    "doc_freq": 5,
    "ttf": 6,
    "term_freq": 1
  },
  "facilit": {
    "doc_freq": 2,
    "ttf": 2,
    "term_freq": 1
  },
  "flexibl": {
    "doc_freq": 4,
    "ttf": 8,
    "term_freq": 5
  },
  "from": {
    "doc_freq": 15,
    "ttf": 30,
    "term_freq": 2
  },
  "further": {
    "doc_freq": 10,
    "ttf": 10,
    "term_freq": 1
  },
  "gener": {
    "doc_freq": 4,
    "ttf": 4,
    "term_freq": 1
  },
  "green": {
    "doc_freq": 3,
    "ttf": 6,
    "term_freq": 1
  },
  "greener": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  },
```

```
"ha": {
  "doc_freq": 6,
  "ttf": 7,
  "term_freq": 1
},
"here": {
  "doc_freq": 3,
  "ttf": 3,
  "term_freq": 1
},
"high": {
  "doc_freq": 9,
  "ttf": 16,
  "term_freq": 1
},
"highli": {
  "doc_freq": 2,
  "ttf": 2,
  "term_freq": 1
},
"holist": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"how": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"howev": {
  "doc_freq": 10,
  "ttf": 11,
  "term_freq": 1
},
"hydrogen": {
  "doc_freq": 5,
  "ttf": 6,
  "term_freq": 1
},
"impact": {
  "doc_freq": 6,
  "ttf": 6,
  "term_freq": 1
},
"improv": {
  "doc_freq": 9,
  "ttf": 11,
  "term_freq": 1
},
"incorpor": {
  "doc_freq": 7,
  "ttf": 7,
  "term_freq": 1
}
```

```
},
"industri": {
  "doc_freq": 6,
  "ttf": 6,
  "term_freq": 1
},
"insight": {
  "doc_freq": 2,
  "ttf": 2,
  "term_freq": 1
},
"integr": {
  "doc_freq": 4,
  "ttf": 5,
  "term_freq": 2
},
"intermedi": {
  "doc_freq": 1,
  "ttf": 2,
  "term_freq": 2
},
"intermitt": {
  "doc_freq": 2,
  "ttf": 3,
  "term_freq": 2
},
"introduc": {
  "doc_freq": 1,
  "ttf": 2,
  "term_freq": 2
},
"investig": {
  "doc_freq": 6,
  "ttf": 6,
  "term_freq": 1
},
"larg": {
  "doc_freq": 3,
  "ttf": 4,
  "term_freq": 2
},
"lead": {
  "doc_freq": 2,
  "ttf": 2,
  "term_freq": 1
},
"learn": {
  "doc_freq": 4,
  "ttf": 4,
  "term_freq": 1
},
"level": {
  "doc_freq": 6,
  "ttf": 7,
```

```
    "term_freq": 1
  },
  "levelis": {
    "doc_freq": 3,
    "ttf": 5,
    "term_freq": 1
  },
  "load": {
    "doc_freq": 2,
    "ttf": 3,
    "term_freq": 2
  },
  "locat": {
    "doc_freq": 6,
    "ttf": 15,
    "term_freq": 2
  },
  "mean": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "meet": {
    "doc_freq": 5,
    "ttf": 5,
    "term_freq": 1
  },
  "meteorolog": {
    "doc_freq": 2,
    "ttf": 2,
    "term_freq": 1
  },
  "methanol": {
    "doc_freq": 2,
    "ttf": 4,
    "term_freq": 2
  },
  "offer": {
    "doc_freq": 5,
    "ttf": 8,
    "term_freq": 2
  },
  "on": {
    "doc_freq": 6,
    "ttf": 7,
    "term_freq": 1
  },
  "oper": {
    "doc_freq": 7,
    "ttf": 9,
    "term_freq": 2
  },
  "optimis": {
    "doc_freq": 5,
```



```
    "ttf": 8,
    "term_freq": 2
  },
  "origin": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "over": {
    "doc_freq": 7,
    "ttf": 8,
    "term_freq": 1
  },
  "overall": {
    "doc_freq": 5,
    "ttf": 6,
    "term_freq": 1
  },
  "pathwai": {
    "doc_freq": 4,
    "ttf": 6,
    "term_freq": 1
  },
  "penetr": {
    "doc_freq": 2,
    "ttf": 4,
    "term_freq": 1
  },
  "perform": {
    "doc_freq": 8,
    "ttf": 13,
    "term_freq": 2
  },
  "possibl": {
    "doc_freq": 6,
    "ttf": 6,
    "term_freq": 1
  },
  "potenti": {
    "doc_freq": 14,
    "ttf": 28,
    "term_freq": 2
  },
  "power": {
    "doc_freq": 4,
    "ttf": 7,
    "term_freq": 2
  },
  "prefer": {
    "doc_freq": 3,
    "ttf": 3,
    "term_freq": 1
  },
  "process": {
```

```
    "doc_freq": 6,
    "ttf": 21,
    "term_freq": 8
  },
  "product": {
    "doc_freq": 15,
    "ttf": 44,
    "term_freq": 7
  },
  "prohibit": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "reduc": {
    "doc_freq": 7,
    "ttf": 9,
    "term_freq": 1
  },
  "renew": {
    "doc_freq": 6,
    "ttf": 24,
    "term_freq": 7
  },
  "respect": {
    "doc_freq": 4,
    "ttf": 8,
    "term_freq": 2
  },
  "reveal": {
    "doc_freq": 3,
    "ttf": 3,
    "term_freq": 1
  },
  "rout": {
    "doc_freq": 5,
    "ttf": 7,
    "term_freq": 1
  },
  "scale": {
    "doc_freq": 6,
    "ttf": 8,
    "term_freq": 2
  },
  "show": {
    "doc_freq": 12,
    "ttf": 13,
    "term_freq": 1
  },
  "size": {
    "doc_freq": 3,
    "ttf": 4,
    "term_freq": 1
  },
  },
```

```
"solar": {
  "doc_freq": 5,
  "ttf": 6,
  "term_freq": 1
},
"sole": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"specifi": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"state": {
  "doc_freq": 2,
  "ttf": 2,
  "term_freq": 1
},
"steadi": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"storag": {
  "doc_freq": 3,
  "ttf": 9,
  "term_freq": 5
},
"studi": {
  "doc_freq": 13,
  "ttf": 18,
  "term_freq": 3
},
"suggest": {
  "doc_freq": 2,
  "ttf": 2,
  "term_freq": 1
},
"suppli": {
  "doc_freq": 13,
  "ttf": 23,
  "term_freq": 1
},
"system": {
  "doc_freq": 13,
  "ttf": 25,
  "term_freq": 1
},
"target": {
  "doc_freq": 10,
  "ttf": 21,
  "term_freq": 1
}
```

```
},
"through": {
  "doc_freq": 11,
  "ttf": 14,
  "term_freq": 3
},
"throughout": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"toward": {
  "doc_freq": 4,
  "ttf": 4,
  "term_freq": 1
},
"two": {
  "doc_freq": 5,
  "ttf": 8,
  "term_freq": 2
},
"unit": {
  "doc_freq": 3,
  "ttf": 5,
  "term_freq": 1
},
"us": {
  "doc_freq": 16,
  "ttf": 32,
  "term_freq": 2
},
"variabl": {
  "doc_freq": 1,
  "ttf": 3,
  "term_freq": 3
},
"via": {
  "doc_freq": 2,
  "ttf": 2,
  "term_freq": 1
},
"viabl": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"we": {
  "doc_freq": 13,
  "ttf": 30,
  "term_freq": 2
},
"when": {
  "doc_freq": 4,
  "ttf": 4,
```

```

    "term_freq": 1
  },
  "which": {
    "doc_freq": 16,
    "ttf": 28,
    "term_freq": 2
  },
  "while": {
    "doc_freq": 5,
    "ttf": 5,
    "term_freq": 1
  },
  "wind": {
    "doc_freq": 4,
    "ttf": 4,
    "term_freq": 1
  },
  "work": {
    "doc_freq": 8,
    "ttf": 8,
    "term_freq": 1
  },
  "year": {
    "doc_freq": 6,
    "ttf": 8,
    "term_freq": 1
  }
}
}
}
}
}

```

Request for TF stats in document/hit 3:

```

GET /zhangqirui_articles/_termvectors/7
{
  "fields": ["abstract"],
  "term_statistics": true,
  "field_statistics": false,
  "positions": false,
  "offsets": false
}

```

Response:

```

{
  "_index": "zhangqirui_articles",
  "_id": "7",
  "_version": 2,

```

```
"found": true,
"took": 2,
"term_vectors": {
  "abstract": {
    "terms": {
      "0": {
        "doc_freq": 1,
        "ttf": 1,
        "term_freq": 1
      },
      "12": {
        "doc_freq": 1,
        "ttf": 2,
        "term_freq": 2
      },
      "23": {
        "doc_freq": 1,
        "ttf": 1,
        "term_freq": 1
      },
      "32": {
        "doc_freq": 1,
        "ttf": 1,
        "term_freq": 1
      },
      "40": {
        "doc_freq": 1,
        "ttf": 1,
        "term_freq": 1
      },
      "45": {
        "doc_freq": 3,
        "ttf": 3,
        "term_freq": 1
      },
      "66": {
        "doc_freq": 1,
        "ttf": 1,
        "term_freq": 1
      },
      "70": {
        "doc_freq": 2,
        "ttf": 2,
        "term_freq": 1
      },
      "74": {
        "doc_freq": 1,
        "ttf": 1,
        "term_freq": 1
      },
      "129": {
        "doc_freq": 1,
        "ttf": 1,
        "term_freq": 1
      }
    }
  }
}
```

```
},
"137": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"192": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"328": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"2030": {
  "doc_freq": 3,
  "ttf": 5,
  "term_freq": 3
},
"2050": {
  "doc_freq": 2,
  "ttf": 4,
  "term_freq": 3
},
"abund": {
  "doc_freq": 4,
  "ttf": 4,
  "term_freq": 1
},
"acceler": {
  "doc_freq": 2,
  "ttf": 2,
  "term_freq": 1
},
"achiev": {
  "doc_freq": 6,
  "ttf": 8,
  "term_freq": 1
},
"across": {
  "doc_freq": 4,
  "ttf": 4,
  "term_freq": 1
},
"ammonia": {
  "doc_freq": 1,
  "ttf": 2,
  "term_freq": 2
},
"arc": {
  "doc_freq": 3,
  "ttf": 3,
```

```
    "term_freq": 1
  },
  "assess": {
    "doc_freq": 6,
    "ttf": 8,
    "term_freq": 1
  },
  "australia": {
    "doc_freq": 1,
    "ttf": 5,
    "term_freq": 5
  },
  "averag": {
    "doc_freq": 4,
    "ttf": 7,
    "term_freq": 2
  },
  "avoid": {
    "doc_freq": 3,
    "ttf": 3,
    "term_freq": 1
  },
  "base": {
    "doc_freq": 14,
    "ttf": 29,
    "term_freq": 5
  },
  "billion": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "binat": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "bonis": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "both": {
    "doc_freq": 7,
    "ttf": 8,
    "term_freq": 1
  },
  "call": {
    "doc_freq": 2,
    "ttf": 2,
    "term_freq": 1
  },
  "capit": {
    "doc_freq": 1,
```



```
    "ttf": 1,
    "term_freq": 1
  },
  "carbon": {
    "doc_freq": 6,
    "ttf": 8,
    "term_freq": 2
  },
  "carrier": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "case": {
    "doc_freq": 5,
    "ttf": 6,
    "term_freq": 1
  },
  "chain": {
    "doc_freq": 5,
    "ttf": 11,
    "term_freq": 5
  },
  "co": {
    "doc_freq": 3,
    "ttf": 4,
    "term_freq": 2
  },
  "co2": {
    "doc_freq": 6,
    "ttf": 23,
    "term_freq": 2
  },
  "configur": {
    "doc_freq": 2,
    "ttf": 2,
    "term_freq": 1
  },
  "consumpt": {
    "doc_freq": 7,
    "ttf": 15,
    "term_freq": 1
  },
  "cost": {
    "doc_freq": 9,
    "ttf": 26,
    "term_freq": 3
  },
  "crack": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "crucial": {
```

```
    "doc_freq": 4,
    "ttf": 4,
    "term_freq": 1
  },
  "current": {
    "doc_freq": 5,
    "ttf": 5,
    "term_freq": 1
  },
  "decad": {
    "doc_freq": 2,
    "ttf": 2,
    "term_freq": 1
  },
  "decar": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "decarbonis": {
    "doc_freq": 5,
    "ttf": 9,
    "term_freq": 2
  },
  "demand": {
    "doc_freq": 5,
    "ttf": 12,
    "term_freq": 1
  },
  "deploi": {
    "doc_freq": 2,
    "ttf": 2,
    "term_freq": 1
  },
  "develop": {
    "doc_freq": 9,
    "ttf": 21,
    "term_freq": 1
  },
  "differenti": {
    "doc_freq": 3,
    "ttf": 3,
    "term_freq": 1
  },
  "diminish": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "direct": {
    "doc_freq": 3,
    "ttf": 6,
    "term_freq": 3
  },
  },
```

```
"disrupt": {
  "doc_freq": 3,
  "ttf": 3,
  "term_freq": 1
},
"domest": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"domin": {
  "doc_freq": 3,
  "ttf": 3,
  "term_freq": 1
},
"drive": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"due": {
  "doc_freq": 8,
  "ttf": 8,
  "term_freq": 1
},
"econom": {
  "doc_freq": 4,
  "ttf": 5,
  "term_freq": 1
},
"economi": {
  "doc_freq": 2,
  "ttf": 2,
  "term_freq": 1
},
"effici": {
  "doc_freq": 6,
  "ttf": 10,
  "term_freq": 3
},
"electr": {
  "doc_freq": 6,
  "ttf": 8,
  "term_freq": 3
},
"enabl": {
  "doc_freq": 3,
  "ttf": 4,
  "term_freq": 2
},
"energet": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
}
```

```
},
"energi": {
  "doc_freq": 12,
  "ttf": 64,
  "term_freq": 12
},
"equival": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"expens": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"export": {
  "doc_freq": 1,
  "ttf": 3,
  "term_freq": 3
},
"flexibl": {
  "doc_freq": 4,
  "ttf": 8,
  "term_freq": 1
},
"focus": {
  "doc_freq": 5,
  "ttf": 5,
  "term_freq": 1
},
"from": {
  "doc_freq": 15,
  "ttf": 30,
  "term_freq": 1
},
"furnac": {
  "doc_freq": 2,
  "ttf": 2,
  "term_freq": 1
},
"further": {
  "doc_freq": 10,
  "ttf": 10,
  "term_freq": 1
},
"global": {
  "doc_freq": 8,
  "ttf": 12,
  "term_freq": 1
},
"green": {
  "doc_freq": 3,
  "ttf": 6,
```

```
    "term_freq": 3
  },
  "highest": {
    "doc_freq": 2,
    "ttf": 2,
    "term_freq": 1
  },
  "hydrogen": {
    "doc_freq": 5,
    "ttf": 6,
    "term_freq": 2
  },
  "hypothesi": {
    "doc_freq": 1,
    "ttf": 2,
    "term_freq": 2
  },
  "i.": {
    "doc_freq": 6,
    "ttf": 7,
    "term_freq": 2
  },
  "immens": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "increas": {
    "doc_freq": 8,
    "ttf": 10,
    "term_freq": 1
  },
  "infrastructur": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "intens": {
    "doc_freq": 5,
    "ttf": 7,
    "term_freq": 1
  },
  "intermitt": {
    "doc_freq": 2,
    "ttf": 3,
    "term_freq": 1
  },
  "invest": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "investig": {
    "doc_freq": 6,
```

```
    "ttf": 6,  
    "term_freq": 1  
  },  
  "iron": {  
    "doc_freq": 2,  
    "ttf": 8,  
    "term_freq": 4  
  },  
  "it": {  
    "doc_freq": 7,  
    "ttf": 8,  
    "term_freq": 1  
  },  
  "japan": {  
    "doc_freq": 1,  
    "ttf": 2,  
    "term_freq": 2  
  },  
  "japanes": {  
    "doc_freq": 1,  
    "ttf": 1,  
    "term_freq": 1  
  },  
  "latter": {  
    "doc_freq": 1,  
    "ttf": 1,  
    "term_freq": 1  
  },  
  "levelis": {  
    "doc_freq": 3,  
    "ttf": 5,  
    "term_freq": 1  
  },  
  "liquefi": {  
    "doc_freq": 1,  
    "ttf": 1,  
    "term_freq": 1  
  },  
  "locat": {  
    "doc_freq": 6,  
    "ttf": 15,  
    "term_freq": 2  
  },  
  "make": {  
    "doc_freq": 2,  
    "ttf": 2,  
    "term_freq": 1  
  },  
  "manufactur": {  
    "doc_freq": 2,  
    "ttf": 9,  
    "term_freq": 1  
  },  
  "meet": {
```

```
    "doc_freq": 5,
    "ttf": 5,
    "term_freq": 1
  },
  "mitig": {
    "doc_freq": 3,
    "ttf": 3,
    "term_freq": 1
  },
  "model": {
    "doc_freq": 13,
    "ttf": 26,
    "term_freq": 1
  },
  "more": {
    "doc_freq": 7,
    "ttf": 10,
    "term_freq": 3
  },
  "mtpa": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "need": {
    "doc_freq": 7,
    "ttf": 11,
    "term_freq": 1
  },
  "offer": {
    "doc_freq": 5,
    "ttf": 8,
    "term_freq": 1
  },
  "onli": {
    "doc_freq": 2,
    "ttf": 2,
    "term_freq": 1
  },
  "oppos": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "or": {
    "doc_freq": 2,
    "ttf": 8,
    "term_freq": 4
  },
  "output": {
    "doc_freq": 2,
    "ttf": 3,
    "term_freq": 1
  },
  },
```

```
"pair": {
  "doc_freq": 1,
  "ttf": 2,
  "term_freq": 2
},
"partnership": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"pilbara": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"plan": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"potenti": {
  "doc_freq": 14,
  "ttf": 28,
  "term_freq": 1
},
"power": {
  "doc_freq": 4,
  "ttf": 7,
  "term_freq": 1
},
"premium": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"prior": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"process": {
  "doc_freq": 6,
  "ttf": 21,
  "term_freq": 2
},
"produc": {
  "doc_freq": 4,
  "ttf": 5,
  "term_freq": 2
},
"product": {
  "doc_freq": 15,
  "ttf": 44,
  "term_freq": 6
```



```
},
"profil": {
  "doc_freq": 2,
  "ttf": 2,
  "term_freq": 1
},
"rang": {
  "doc_freq": 3,
  "ttf": 3,
  "term_freq": 1
},
"realis": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"reduc": {
  "doc_freq": 7,
  "ttf": 9,
  "term_freq": 1
},
"reduct": {
  "doc_freq": 7,
  "ttf": 11,
  "term_freq": 3
},
"region": {
  "doc_freq": 4,
  "ttf": 13,
  "term_freq": 4
},
"remot": {
  "doc_freq": 1,
  "ttf": 1,
  "term_freq": 1
},
"renew": {
  "doc_freq": 6,
  "ttf": 24,
  "term_freq": 4
},
"requir": {
  "doc_freq": 5,
  "ttf": 12,
  "term_freq": 3
},
"resourc": {
  "doc_freq": 5,
  "ttf": 17,
  "term_freq": 3
},
"respect": {
  "doc_freq": 4,
  "ttf": 8,
```

```
    "term_freq": 1
  },
  "result": {
    "doc_freq": 15,
    "ttf": 23,
    "term_freq": 1
  },
  "rout": {
    "doc_freq": 5,
    "ttf": 7,
    "term_freq": 1
  },
  "scale": {
    "doc_freq": 6,
    "ttf": 8,
    "term_freq": 2
  },
  "set": {
    "doc_freq": 5,
    "ttf": 7,
    "term_freq": 1
  },
  "ship": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "simul": {
    "doc_freq": 4,
    "ttf": 8,
    "term_freq": 1
  },
  "solar": {
    "doc_freq": 5,
    "ttf": 6,
    "term_freq": 2
  },
  "sourc": {
    "doc_freq": 3,
    "ttf": 4,
    "term_freq": 1
  },
  "spatial": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "steel": {
    "doc_freq": 2,
    "ttf": 13,
    "term_freq": 8
  },
  "strateg": {
    "doc_freq": 1,
```

```
    "ttf": 1,
    "term_freq": 1
  },
  "structur": {
    "doc_freq": 3,
    "ttf": 3,
    "term_freq": 1
  },
  "studi": {
    "doc_freq": 13,
    "ttf": 18,
    "term_freq": 2
  },
  "suppli": {
    "doc_freq": 13,
    "ttf": 23,
    "term_freq": 5
  },
  "system": {
    "doc_freq": 13,
    "ttf": 25,
    "term_freq": 1
  },
  "t": {
    "doc_freq": 4,
    "ttf": 10,
    "term_freq": 2
  },
  "tax": {
    "doc_freq": 1,
    "ttf": 2,
    "term_freq": 2
  },
  "techno": {
    "doc_freq": 2,
    "ttf": 2,
    "term_freq": 1
  },
  "technolog": {
    "doc_freq": 5,
    "ttf": 7,
    "term_freq": 1
  },
  "test": {
    "doc_freq": 2,
    "ttf": 2,
    "term_freq": 1
  },
  "than": {
    "doc_freq": 6,
    "ttf": 10,
    "term_freq": 2
  },
  "through": {
```

```
    "doc_freq": 11,
    "ttf": 14,
    "term_freq": 1
  },
  "timelin": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "total": {
    "doc_freq": 3,
    "ttf": 4,
    "term_freq": 1
  },
  "tradit": {
    "doc_freq": 4,
    "ttf": 4,
    "term_freq": 1
  },
  "two": {
    "doc_freq": 5,
    "ttf": 8,
    "term_freq": 1
  },
  "uniqu": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "us": {
    "doc_freq": 16,
    "ttf": 32,
    "term_freq": 1
  },
  "valid": {
    "doc_freq": 7,
    "ttf": 8,
    "term_freq": 1
  },
  "vector": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "vessel": {
    "doc_freq": 1,
    "ttf": 1,
    "term_freq": 1
  },
  "well": {
    "doc_freq": 5,
    "ttf": 6,
    "term_freq": 1
  },
  },
```

```
      "were": {
        "doc_freq": 10,
        "ttf": 16,
        "term_freq": 2
      },
      "western": {
        "doc_freq": 1,
        "ttf": 1,
        "term_freq": 1
      },
      "when": {
        "doc_freq": 4,
        "ttf": 4,
        "term_freq": 1
      },
      "which": {
        "doc_freq": 16,
        "ttf": 28,
        "term_freq": 2
      },
      "wholli": {
        "doc_freq": 1,
        "ttf": 1,
        "term_freq": 1
      },
      "would": {
        "doc_freq": 2,
        "ttf": 4,
        "term_freq": 3
      }
    }
  }
}
```

Query keyword TFs and DFs in the top three hits:

Doc ID	Term 1 TF	Term 1 DF	Term 2 TF	Term 2 DF	Term 3 TF	Term 3 DF
15	9	6	3	4	1	13
18	7	6	2	4	1	13
7	4	6	1	4	5	13

B. TF*IDF (1 point)

TF*IDF weights:

Doc ID	Term 1	Term 2	Term 3	SUM
15	4.7	2.1	0.2	7

Doc ID	Term 1	Term 2	Term 3	SUM
18	3.7	1.4	0.2	5.3
7	2.1	0.7	0.2	3

Rank the documents using SUM of TF*IDF.
document15, document18, document7

Are they ranked in the same order as in ElasticSearch results?
Yes.

C. Cosine Similarity (1 point)

Query IDF (\$1\times IDF\$) weights:

Doc ID	Term 1	Term 2	Term 3
15	0.5	0.7	0.2
18	0.5	0.7	0.2
7	0.5	0.7	0.2

Cosine similarities:

1. $\$Cosine(Q, doc1) = \frac{0.5 \cdot 4.7 + 0.7 \cdot 2.1 + 0.2 \cdot 0.2}{\sqrt{(0.5^2 + 0.7^2 + 0.2^2)(4.7^2 + 2.1^2 + 0.2^2)}} = 3.86/4.55 = 0.85 \$$
2. $\$Cosine(Q, doc2) = \frac{0.5 \cdot 3.7 + 0.7 \cdot 1.4 + 0.2 \cdot 0.2}{\sqrt{(0.5^2 + 0.7^2 + 0.2^2)(3.7^2 + 1.4^2 + 0.2^2)}} = 3.89/5.84 = 0.67 \$$
3. $\$Cosine(Q, doc3) = \frac{0.5 \cdot 2.1 + 0.7 \cdot 0.7 + 0.2 \cdot 0.2}{\sqrt{(0.5^2 + 0.7^2 + 0.2^2)(2.1^2 + 0.7^2 + 0.2^2)}} = 1.58/1.96 = 0.81 \$$

Rank the documents using Cosine scores.
document13, document7, document18

Are they ranked in the same order as in TF*IDF SUM or ElasticSearch results?
No.