Web Crawler Report

1. How many unique pages did you find?
"unique_url_count": 5919,

2. What is the longest page in terms of the number of words?
"longest_page": {
    "url": "https://cml.ics.uci.edu/category/aiml",
    "word_count": 16340
  },

3. What are the 50 most common words?
},
  "top_words": [
   [
     "research",
     17312
   ],
   [
     "computer",
     11867
   ],
   [
     "data",
     11808
   ],
   [
     "ics",
     11639
   ],
   [
     "information",
     11298
   ],
   [
     "uci",
     10353
   ],
   [
     "ramesh",
     10262
   ],
   [

```
    "will",
    9843
  ],
  [
    "students",
    9799
  ],
  [
    "can",
    9311
  ],
  [
    "software",
    8788
  ],
  [
    "student",
    8512
  ],
  [
    "events",
    7923
  ],
  [
    "science",
    7818
  ],
  [
    "university",
    7622
  ],
  [
    "us",
    7600
  ],
  [
    "learning",
    7084
  ],
  [
    "news",
    6937
  ],
  [
```

```
    "one",
    6803
  ],
  [
    "current",
    6793
  ],
  [
    "may",
    6785
  ],
  [
    "search",
    6649
  ],
  [
    "time",
    6455
  ],
  [
    "computing",
    6299
  ],
  [
    "people",
    6234
  ],
  [
    "projects",
    6140
  ],
  [
    "irvine",
    6130
  ],
  [
    "past",
    6077
  ],
  [
    "systems",
    5888
  ],
  [
```

```
      "new",
      5857
    ],
    [
      "engineering",
      5804
    ],
    [
      "graduate",
      5797
    ],
    [
      "courses",
      5761
    ],
    [
      "contact",
      5463
    ],
    [
      "undergraduate",
      5441
    ],
    [
      "jain",
      5433
    ],
    [
      "informatics",
      5341
    ],
    [
      "read",
      5101
    ],
    [
      "wics",
      5048
    ],
    [
      "project",
      5003
    ],
    [
```

    "books",
    4975
  ],
  [
    "also",
    4848
  ],
  [
    "using",
    4780
  ],
  [
    "code",
    4746
  ],
  [
    "program",
    4726
  ],
  [
    "uc",
    4677
  ],
  [
    "pm",
    4446
  ],
  [
    "vision",
    4361
  ],
  [
    "design",
    4342
  ],
  [
    "support",
    4311
  ],
  [
    "department",
    4256
  ],
  [

```
    "community",
    4220
  ],
  [
    "faculty",
    4211
  ],
  [
    "use",
    4211
  ],
  [
    "week",
    4083
  ],
  [
    "machine",
    4061
  ],
  [
    "work",
    4004
  ],
  [
    "home",
    3936
  ],
  [
    "based",
    3907
  ],
  [
    "edu",
    3864
  ],
  [
    "management",
    3855
  ],
  [
    "companies",
    3809
  ],
  [
```

```
      "social",
      3767
    ],
    [
      "ph",
      3758
    ],
    [
      "first",
      3752
    ],
    [
      "applications",
      3730
    ],
    [
      "conference",
      3633
    ],
    [
      "october",
      3593
    ],
    [
      "school",
      3545
    ],
    [
      "two",
      3504
    ],
    [
      "web",
      3503
    ],
    [
      "june",
      3490
    ],
    [
      "technology",
      3484
    ],
    [
```

    "many",
    3468
  ],
  [
    "eppstein",
    3452
  ],
  [
    "paper",
    3413
  ],
  [
    "like",
    3383
  ],
  [
    "system",
    3340
  ],
  [
    "theme",
    3332
  ],
  [
    "favorite",
    3331
  ],
  [
    "blogs",
    3304
  ],
  [
    "publications",
    3302
  ],
  [
    "august",
    3297
  ],
  [
    "algorithms",
    3262
  ],
  [

```
    "future",
    3243
  ],
  [
    "alumni",
    3236
  ],
  [
    "september",
    3225
  ],
  [
    "process",
    3167
  ],
  [
    "march",
    3162
  ],
  [
    "july",
    3159
  ],
  [
    "april",
    3157
  ],
  [
    "november",
    3143
  ],
  [
    "models",
    3117
  ],
  [
    "health",
    3068
  ],
  [
    "navigation",
    3055
  ],
  [
```

```
    "experiential",
    3047
  ],
  [
    "set",
    3043
  ],
  [
    "january",
    3017
  ],
  [
    "see",
    3009
  ],
  [
    "get",
    2978
  ]
],
```

4. How many subdomains did you find in the uci.edu domain?

```
"subdomains": {
    "accessibility.ics.uci.edu": 5,
    "acoi.ics.uci.edu": 72,
    "archive-beta.ics.uci.edu": 7,
    "archive.ics.uci.edu": 142,
    "asterix.ics.uci.edu": 5,
    "betapro.proteomics.ics.uci.edu": 2,
    "capstone.cs.uci.edu": 8,
    "cdb.ics.uci.edu": 28,
    "cert.ics.uci.edu": 6,
    "chenli.ics.uci.edu": 8,
    "cloudberry.ics.uci.edu": 24,
    "cml.ics.uci.edu": 62,
    "code.ics.uci.edu": 13,
    "computableplant.ics.uci.edu": 31,
    "courselisting.ics.uci.edu": 3,
    "create.ics.uci.edu": 5,
    "cs.ics.uci.edu": 10,
    "cs.uci.edu": 1,
    "cwicsocal18.ics.uci.edu": 7,
```

```
"cyberclub.ics.uci.edu": 9,
"cybert.ics.uci.edu": 23,
"dgillen.ics.uci.edu": 9,
"ds4all.ics.uci.edu": 1,
"duttgroup.ics.uci.edu": 29,
"edgelab.ics.uci.edu": 6,
"emj.ics.uci.edu": 27,
"evoke.ics.uci.edu": 1,
"flamingo.ics.uci.edu": 7,
"fr.ics.uci.edu": 3,
"futurehealth.ics.uci.edu": 131,
"grape.ics.uci.edu": 82,
"graphics.ics.uci.edu": 4,
"graphmod.ics.uci.edu": 1,
"hai.ics.uci.edu": 4,
"hobbes.ics.uci.edu": 9,
"hpi.ics.uci.edu": 5,
"hub.ics.uci.edu": 1,
"icde2023.ics.uci.edu": 37,
"ics.uci.edu": 209,
"industryshowcase.ics.uci.edu": 17,
"informatics.ics.uci.edu": 1,
"informatics.uci.edu": 2,
"insite.ics.uci.edu": 8,
"ipubmed.ics.uci.edu": 1,
"isg.ics.uci.edu": 234,
"jgarcia.ics.uci.edu": 37,
"julia-hub.ics.uci.edu": 1,
"luci.ics.uci.edu": 4,
"mailman.ics.uci.edu": 11,
"mcs.ics.uci.edu": 10,
"mdogucu.ics.uci.edu": 1,
"mds.ics.uci.edu": 4,
"mhcid.ics.uci.edu": 16,
"mlphysics.ics.uci.edu": 2,
"motifmap.ics.uci.edu": 1,
"mswe.ics.uci.edu": 7,
"mupro.proteomics.ics.uci.edu": 2,
"nalini.ics.uci.edu": 4,
"ngs.ics.uci.edu": 1571,
"oai.ics.uci.edu": 5,
"psearch.ics.uci.edu": 1,
"redmiles.ics.uci.edu": 1,
"riscit.ics.uci.edu": 2,
```

    "scratch.proteomics.ics.uci.edu": 3,
    "seal.ics.uci.edu": 41,
    "selectpro.proteomics.ics.uci.edu": 4,
    "sherlock.ics.uci.edu": 6,
    "speedtest.ics.uci.edu": 1,
    "stairs.ics.uci.edu": 4,
    "stat.uci.edu": 1,
    "statconsulting.ics.uci.edu": 1,
    "student-council.ics.uci.edu": 1,
    "summeracademy.ics.uci.edu": 5,
    "swiki.ics.uci.edu": 101,
    "tad.ics.uci.edu": 1,
    "tastier.ics.uci.edu": 1,
    "tutoring.ics.uci.edu": 5,
    "unite.ics.uci.edu": 10,
    "vision.ics.uci.edu": 189,
    "wics.ics.uci.edu": 649,
    "wiki.ics.uci.edu": 42,
    "www-db.ics.uci.edu": 19,
    "www.ics.uci.edu": 1408,
    "www.informatics.ics.uci.edu": 1,
    "www.informatics.uci.edu": 361,
    "www.stat.uci.edu": 82,
    "xtune.ics.uci.edu": 3
  }
}