Overview of the NTCIR-15 Data Search Task

Makoto P. Kato
University of Tsukuba
mpkato@acm.org

Hiroaki Ohshima
University of Hyogo
ohshima@ai.u-hyogo.ac.jp

Ying-Hsang Liu
University of Southern Denmark
yingliu@sdu.dk

Hsin-Liang Chen
Missouri University of Science and Technology
chenhs@mst.edu

ABSTRACT

NTCIR-15 Data Search is a shared task on ad-hoc retrieval for governmental statistical data. The first round of Data Search focuses on the retrieval of a statistical data collection published by the Japanese government (e-Stat), and one published by the US government (Data.gov). This paper introduces the task definition, test collection, and evaluation methodology of NTCIR-15 Data Search. This round of Data Search attracted six research groups, from which we received 17 submissions for the Japanese subtask, and 37 submissions for the English subtask. The evaluation results of these runs are presented and discussed in this paper.

1 INTRODUCTION

The open data movement is now being accelerated by the expectations for open science and citizen science. It is said that researchers all over the world could collaborate on world-wide problems and citizens also could participate in research activities if various kinds of data were publicly available. The government of each country strongly encourages the open data movement and has launched open-data government initiatives such as Data.gov[1] in the United States, Data.gov.uk in the United Kingdom, and e-Stat[2] in Japan. Besides the governmental portals, there are also thousands of data repositories on the Web [6].

The growth of the open data movement has naturally motivated researchers and industries to develop search engines for the open data scattered on the Web. Google launched Google Dataset Search as public beta in September 2018 [7], and some researchers have started to discuss potential research topics of data search [2]. Although there have been several attempts for understanding and developing data search, neither a benchmark nor an evaluation campaign on data search has not been proposed yet.

Therefore, following rapidly increasing demands and interests in data search, we propose a pilot task on data search, *Data Search*, at NTCIR-15. The first round of Data Search focuses on a standard ad-hoc retrieval task for data. To be more specific, we target the retrieval of a statistical data collection published by the Japanese government (e-Stat), and one published by the US government (Data.gov). We provide a set of topics derived from questions in a community question-answering service, queries generated by users based on the topics, and relevance grades for a subset of the statistical data.

Through the Data Search task, we expect advances in the following areas:

Query understanding for data search According to the query log analysis of open data portals [3] and our observations on candidate queries for Data Search, queries for data search include more geographical, temporal, and numerical keywords than those for Web search. Furthermore, as suggested by Koesten *et al.* [4], the goal of data search can be diverse, *e.g.* time series analysis and summarization. Thus, queries for data search need a dedicated interpretation technique and to be studied for a better retrieval performance.

Data understanding for data search Metadata of data usually include the name, short description, category, and date. They are used for indexing data, but are not always sufficiently informative for data search. Data are often released in Excel, CSV, XML, and PDF formats, and structured in tables for many cases. They could be potentially used with metadata to enrich the index for data search, while interpreting data on the Web is a still challenging problem.

Retrieval models for data search Data and their metadata contain a lot of entities such as locations or products, temporal expressions, and numerical expressions. Hence, retrieval models for entity or temporal information could be effective in data search as well. Numerical expressions might require a new model for better rankings. Moreover, retrieval models might need to be adaptive depending on the goal of data search.

In the remainder of this paper, we briefly explain related work on data search in Section 2, introduce the task, resources, and evaluation methodology in Section 3, and finally discuss evaluation results in Section 4.

2 RELATED WORK

The initial research on data search mainly focused on how people search for data. Kacprzak *et al.* investigated queries for four national open data portals [3]. Their analysis revealed that (1) 90% of queries are 1-3 words queries. The average length is 2.03, (2) location keywords, temporal keywords, file and dataset types, and numbers are included in 5-8% of queries, and (3) there are a small number of question type queries (less than 1%). Koesten *et al.* studied the information seeking behavior in data search [4]. Based on interviews to several types of data users, they developed a taxonomy of activities with data (Process-oriented tasks vs. Goal-oriented tasks, and five major activities: linking, time series analysis, summarizing, presenting, and exporting), identified major relevance criteria such as relevance, usability, and quality, and found a typical workflow after finding relevant data (*e.g.* looking at headers, looking for obvious errors, and looking at summarizing statistics).

[1]https://www.data.gov/
[2]https://www.e-stat.go.jp/

Table 1: Statistics of the test collections.

Subtask	Resource	#
Japanese		
	Datasets	1,338,402
	Data files	1,338,402
	Training queries	96
	Training qrels	2,035
	Test queries	96
	Test qrels	5,719
English		
	Datasets	46,615
	Data files	92,930
	Training queries	96
	Training qrels	2,008
	Test queries	96
	Test qrels	6,240

There are several research directions that do not explicitly refer to data search, but are potentially related. Table search is one of the most related work to data search, since data are often represented in a form of table. Zhang and Balog tackled a problem of ad-hoc table retrieval [12], and proposed table-specific features effective for table retrieval and similarity measures for query-table matching. Yakout *et al.* defined three operations for augmenting tables, namely, augmentation by attribute name, augmentation by example, and discovery of important attributes [10]. Table explanation is also a related topic to data search, since it can potentially be used for enriching index or snippet generation in data search [1, 5, 8, 9].

3 METHODOLOGY

This section introduces the methodology of Data Search, including the task, topics, queries, data collections from which data are retrieved, relevance judgments, and the evaluation methodology.

3.1 Task

The task of Data Search is almost the same as standard ad-hoc retrieval tasks, and is defined as follows: Given a query for data search, a system is expected to return a ranked list of *datasets* (precisely defined in the next subsection).

We have two subtasks in the Data Search task, namely, Japanese subtask and English subtask. The e-Stat data are used in the Japanese subtask, while the Data.gov data are used in the English subtask.

In NTCIR-15 Data Search, each team was allowed to submit up to 10 runs. Runs should be generated automatically.

3.2 Resources

The Data Search task provides the following resources[3]. The statistics of the test collections are shown in Table 1.

[3] Available at https://ntcir.datasearch.jp/

3.2.1 Topics. Topics were derived from 3,218 question-answers pairs including links to the Japanese government open data portal, e-Stat, which were crawled from a Japanese community question-answering service, Yahoo! Chiebukuro[4]. We manually examined each question and extracted 192 questions that indicate information needs for data search. Some examples of the topics are shown in Table 2. For the English subtask, we manually translate Japanese topics into English ones with Japanese-specific named entities replaced with corresponding US-specific named entities. For example, "Tokyo" was replaced with "New York", and "shrines" was replaced with "churches". Half of the topics were used as training topics, while the others were used as test topics.

3.2.2 Queries. Since it is not obvious how topics can be translated into queries, and we are also interested in query formulation for data search, we gathered queries for data search by using crowd-sourcing services. A Japanese crowd-sourcing service, Lancers[5], was used for Japanese topics, while Amazon Mechanical Turk[6] was used for English topics. Ten workers were given a topic and asked to input a query for data search. The exact instruction we provided is shown below:

> You are given a request or a question from someone who wants to get certain information or an answer to the question. Please type some keywords for a web search to provide her/his desired information or answer.

We then selected the most representative query for each topic as follows. For each query, we compute the cross entropy between the language models of the topic and query:

$$H(t,q) = -\sum_{w \in q} P(w|q) \log P(w|t) \tag{1}$$

where $P(w|t)$ is estimated by the frequency of w in the queries given for t, and $P(w|q)$ is estimated by the frequency of w in q. The low entropy indicates the closeness of the two language models, which suggests that the query language model is close to that for the entire query set for a topic. This can be considered as the representativeness in the given topic. Thus, we chose query q that minimizes $H(t,q)$ as the most representative query for topic q. Some examples of the queries are shown in Table 2, together with their topics.

3.2.3 Data. We crawled around 1.3 millions of pages in e-Stat and 0.2 millions of pages in data.gov. The e-Stat pages only describe the metadata of a single data file, while the data.gov pages summarize the metadata of multiple related data files. Therefore, we define a *dataset* as a pair of metadata and a set of data files, and use the dataset as a unit of retrieval in our search task. In this round of the Data Search task, for convenience of data processing, we restrict the type of data files to Excel (i.e. xls and xlsx), CSV, and PDF files for e-Stat data files, and Excel, CSV, PDF, XML, JSON, RDF, and text files for data.gov data files. To increase the availability of the datasets, we used only the datasets allowing redistribution and modification. All the datasets in e-Stats are distributed under a license compatible

[4] https://chiebukuro.yahoo.co.jp/
[5] https://www.lancers.jp/
[6] https://www.mturk.com/

Table 2: Examples of the topics and queries.

Topic ID	Topic	Query
DS1-E-0001	Do people in the East Coast dislike oysters?	oysters dislike east coast
DS1-E-0004	I am looking for evidences of domestic self-sufficiency rate of salt	domestic self salt rate.
DS1-E-0007	Are there many people who can't drive large trailers?	people can't drive large trailers
DS1-E-0009	How many people have a second house?	many people second house
DS1-E-0014	Which city has a population of about 300,000?	city population 300,000

to CC BY[7], which allows redistribution and modification. For the data.gov datasets, we used only the datasets distributed under U.S. Government Work, CC BY, etc. The statistics of the datasets can be found in Table 1.

e-Stat provides diverse kinds of statistical data on weather, population, industry, energy, transportation, education, science, government, judiciary, social security, and so on. The metadata consist of the name, ID, short description, category, publishing organization, survey date, and release date.

data.gov is a portal site of the U.S. Government's open data on agriculture, climate, ecosystems, energy, local government, maritime, ocean, and older adults health. Similar but a little more detailed metadata are given to a set of data files. An example of the metadata is shown in Figure 1.

3.2.4 Relevance Judgments. In addition to the resources described above, we provided relevance scores for some topic-dataset pairs so that participants can evaluate or train their system. We developed several standard baseline systems such as BM25, LM, and BM25+RM3, pooled the top-ranked results for the training queries, and evaluated the relevance grade by crowd-sourcing services. These qrels were distributed to the participants together with the test collections. The details of the relevance judgments are explained in the next subsection.

All the baseline systems were implemented by Anserini [11], and were also introduced to the participants[8].

3.3 Evaluation

The evaluation of Data Search is almost the same as standard ad-hoc retrieval evaluation. For both of the training and test queries, we pooled the top 10 documents of each system for each query. The same crowd-sourcing services as those used for the query generation were used for relevance judgments, namely, Lancers for the Japanese subtask and Amazon Mechanical Turk for the English subtask. Each topic-dataset pair was evaluated at a three-point scale (0: irrelevant, 1: partially relevant, and 2: highly relevant). The exact instruction we provided is:

- Please judge how useful a DATASET of a webpage is for answering a given REQUEST.
- Please carefully read a given REQUEST, visit a webpage describing a DATASET, and give a usefulness score (0, 1, or 2) to each of the datasets.

Rules


```
{
    "id": "0063664a-d0d7-4ce2-9462-0463a89fc274",
    "url": "https://catalog.data.gov/dataset/0063664a-
        d0d7-4ce2-9462-0463a89fc274",
    "attribution": "CRED REA Fish Team Stationary Point
        Count Surveys at Sarigan, Marianas Archipelago,
        2005 (https://catalog.data.gov/dataset/0063664a-
        d0d7-4ce2-9462-0463a89fc274) is licensed under U
        .S. Government Work (http://www.usa.gov/
        publicdomain/label/1.0/)"
    "title": "CRED REA Fish Team Stationary Point Count
        Surveys at Sarigan, Marianas Archipelago, 2005",
    "description": "Stationary Point Counts at 4 stations
         at each survey site were surveyed as part of
        Rapid Ecological Assessments (REA) conducted at
        3 sites around Sarigan in the Marianas
        Archipelago (MA) during 3 September - 1 October
        2005 in the NOAA Oscar Elton Sette (OES 0511)
        Reef Assessment and Monitoring Program (RAMP)
        Cruise. Raw survey data included species level
        abundance estimates.",
    "data": [
        {
            "data_format": "excel",
            "data_organization": "National Oceanic and
                Atmospheric Administration, Department
                of Commerce",
            "data_url": "https://data.nodc.noaa.gov/coris
                /data/NOAA/nmfs/pifsc/cred/REAFish/
                CNMI_2005/CRED_REA_FISH_SAIPAN_2005.xls"
                ,
            "data_filename": "CRED_REA_FISH_SAIPAN_2005.
                xls"
        }
    ],
    "data_fields": {
        "Resource Type": "Dataset",
        "Metadata Date": "June 20, 2018",
        "Metadata Created Date": "February 7, 2018",
        "Metadata Updated Date": "February 27, 2019",
        ...
        "metadata_sources": [
            "https://catalog.data.gov/harvest/object/
                fc5a39b7-4c9f-49b8-af95-2812d9b3264c"
        ]
    }
}
```


Figure 1: Example of metadata of an English dataset.

(1) Carefully read a REQUEST (Note: this page contains a few types of requests.)
(2) Make sure that you visit a webpage that describes a DATASET, and judge how useful the DATASET is for answering the REQUEST.
(3) Usefulness score is defined as:

[7] https://creativecommons.org/licenses/by/4.0/legalcode
[8] Available at https://github.com/mpkato/ntcir-datasearch

- 0: (Useless) The DATASET is not useful to answer the REQUEST at all, or was not accessible for some reasons.
- 1: (Partially useful) The DATASET is useful to partially answer the REQUEST, but cannot fully answer the REQUEST.
- 2: (Highly useful) The DATASET is useful to fully answer the REQUEST.

For the relevance judgments for training topics, we assigned five workers to each topic-dataset pair and removed the highest and lowest scores for excluding outliers. To ensure the quality of the assessments, we showed exactly the same topic-document pairs and measured the consistency of the assessments. If over 25% of answers for these topic-document pairs were inconsistent, we excluded such assessors in the evaluation. The inter-rater agreement measured by Krippendorff's α is 0.736 for the Japanese subtask, and 0.344 for the English subtask.

Since we found a low agreement for the English subtask, we updated the crowd-sourcing setting for test topics. We selected topic-dataset pairs for which relevance scores were consistent. More precisely, they are considered consistent if the average score of *five* scores is 1.8 or higher, or 0.2 or lower. Those topic-dataset pairs were used as *gold* data for measuring the performance of each worker. In the evaluation for the test queries, 10% of topic-dataset pairs were gold data. We banned workers who conducted over 30 judgments and made errors for over 30% of gold data. Moreover, we used an option "Require that Workers be Masters to do your tasks" in Amazon Mechanical Turk and found that this significantly increase the quality of the judgments. Five assessors were assigned for each topic-dataset pair for the Japanese subtask, while three assessors were assigned for the English subtask. Krippendorff's α is 0.478 for the Japanese subtask and 0.438 for the English subtask.

Standard evaluation metrics for ad-hoc retrieval tasks, nDCG, ERR, and Q-measure, were used in NTCIR-15 Data Search. nDCG@10 was used as the primary metric of our task. NTCIREVAL was used for computing the effectiveness scores[9].

4 EVALUATION RESULTS

NTCIR-15 Data Search attracted six research groups including two organizer teams that provided baseline runs. There were 17 submissions for the Japanese subtask, and 37 submissions for the English subtask in the NTCIR-15 Data Search task. All the submitted runs are listed in Table 3. Each run was named "[GROUP_ID]-[LANGUAGE]-[PRIORITY]" where "[GROUP_ID]" is a group ID, "[LANGUAGE]" is either "J" (Japanese subtask) or "E" (English subtask), and "[PRIORITY]" is an integer between 1 and 10, indicating which runs should be prioritized in the pooling for relevance assessments. Each run file was required to include a system description, which is also shown in the table.

Tables 4 and 5 show the evaluation results of Japanese and English subtask runs, respectively. Runs are sorted by nDCG@10, which is our primary evaluation metric.

[9]http://research.nii.ac.jp/ntcir/tools/ntcireval-en.html

5 CONCLUSIONS

This paper introduced the task definition, test collection, and evaluation methodology of NTCIR-15 Data Search. This round of Data Search attracted six research groups, from which we received 17 submissions for the Japanese subtask, and 37 submissions for the English subtask.

REFERENCES

[1] Junwei Bao, Duyu Tang, Nan Duan, Zhao Yan, Yuanhua Lv, Ming Zhou, and Tiejun Zhao. 2018. Table-to-text: Describing table region with natural language. In *AAAI*. 5020–5027.

[2] Adriane Chapman, Elena Simperl, Laura Koesten, George Konstantinidis, Luis Daniel Ibáñez-Gonzalez, Emilia Kacprzak, and Paul T. Groth. 2019. Dataset search: a survey. *CoRR* abs/1901.00735 (2019). http://arxiv.org/abs/1901.00735

[3] Emilia Kacprzak, Laura M Koesten, Luis-Daniel Ibáñez, Elena Simperl, and Jeni Tennison. 2017. A query log analysis of dataset search. In *International Conference on Web Engineering*. 429–436.

[4] Laura M Koesten, Emilia Kacprzak, Jenifer FA Tennison, and Elena Simperl. 2017. The trials and tribulations of working with structured data:-a study on information seeking behaviour. In *CHI*. 1277–1289.

[5] Rémi Lebret, David Grangier, and Michael Auli. 2016. Neural Text Generation from Structured Data with Application to the Biography Domain. In *EMNLP*. 1203–1213.

[6] D-Lib Magazine. 2017. The landscape of research data repositories in 2015: A re3data analysis. *D-Lib Magazine* 23, 3/4 (2017).

[7] Natasha Noy, Matthew Burgess, and Dan Brickley. 2019. Google Dataset Search: Building a search engine for datasets in an open Web ecosystem. In *WebConf*. 1365–1375.

[8] Lei Sha, Lili Mou, Tianyu Liu, Pascal Poupart, Sujian Li, Baobao Chang, and Zhifang Sui. 2018. Order-planning neural text generation from structured data. In *AAAI*. 5414–5421.

[9] Sam Wiseman, Stuart Shieber, and Alexander Rush. 2017. Challenges in Data-to-Document Generation. In *EMNLP*. 2253–2263.

[10] Mohamed Yakout, Kris Ganjam, Kaushik Chakrabarti, and Surajit Chaudhuri. 2012. Infogather: entity augmentation and attribute discovery by holistic matching with web tables. In *SIGMOD*. 97–108.

[11] Peilin Yang, Hui Fang, and Jimmy Lin. 2018. Anserini: Reproducible Ranking Baselines Using Lucene. *Journal of Data and Information Quality* 10, 4, Article 16 (Oct. 2018), 20 pages. https://doi.org/10.1145/3239571

[12] Shuo Zhang and Krisztian Balog. 2018. Ad Hoc Table Retrieval Using Semantic Similarity. In *WWW*. 1553–1562.

Table 3: Runs submitted to the NTCIR-15 Data Search task.

Japanese subtask		
Group ID	Run name	Description
KSU	KSU-J-1	category search, QA categories and BM25 and table headers
KSU	KSU-J-3	Birch and table headers
KSU	KSU-J-5	category search, QA categories and BM25
KSU	KSU-J-7	Birch
ORGJ	ORGJ-J-1	ja-bm25prf+bm25
ORGJ	ORGJ-J-2	ja-bm25
ORGJ	ORGJ-J-3	ja-bm25.accurate
ORGJ	ORGJ-J-4	ja-sdm+qld
ORGJ	ORGJ-J-5	ja-rm3+bm25
ORGJ	ORGJ-J-6	ja-qld
ORGJ	ORGJ-J-7	ja-sdm+bm25
ORGJ	ORGJ-J-8	ja-rm3+qld
uhai	uhai-J-6	Query Fixing + L2R + Bert
uhai	uhai-J-7	Query Fixing + L2R
uhai	uhai-J-8	L2R
uhai	uhai-J-9	L2R + Bert
uhai	uhai-J-10	Query Fixing + bm25

English subtask		
Group ID	Run name	Description
KSU	KSU-E-2	category search, QA categories, BM25 and table headers
KSU	KSU-E-4	Birch and table headers
KSU	KSU-E-6	category search, QA categories and BM25
KSU	KSU-E-8	Birch
NIITableLinker	NIITableLinker-E-1	BM25 [fine-tune]
NIITableLinker	NIITableLinker-E-2	BM25+PRF [default]
NIITableLinker	NIITableLinker-E-3	BM25+PRF [fine-tune]
NIITableLinker	NIITableLinker-E-4	R2+BERT
NIITableLinker	NIITableLinker-E-5	R3+BERT
NIITableLinker	NIITableLinker-E-6	Entity + Noun phrase + BM25+PRF
NIITableLinker	NIITableLinker-E-7	DATE LOC
NIITableLinker	NIITableLinker-E-8	metadata attributes + BM25+PRF
NIITableLinker	NIITableLinker-E-9	cluster
NIITableLinker	NIITableLinker-E-10	R3+BERT+Top100
ORGE	ORGE-E-1	en-bm25prf+bm25
ORGE	ORGE-E-2	en-bm25
ORGE	ORGE-E-3	en-bm25.accurate
ORGE	ORGE-E-4	en-sdm+qld
ORGE	ORGE-E-5	en-rm3+bm25
ORGE	ORGE-E-6	en-qld
ORGE	ORGE-E-7	en-sdm+bm25
ORGE	ORGE-E-8	en-rm3+qld
STIS	STIS-E-1	RM3+BM25 AND FINETUNED BERT BERT-BASE-UNCASED
STIS	STIS-E-2	RM3+BM25 AND FINETUNED BERT BERT-BASE-UNCASED
STIS	STIS-E-3	RM3+BM25 AND FINETUNED BERT BERT-LARGE-UNCASED
STIS	STIS-E-4	RM3+BM25 AND FINETUNED BERT BERT-LARGE-UNCASED
STIS	STIS-E-5	RM3+BM25 AND FINETUNED ROBERTA ROBERTA-BASE
STIS	STIS-E-6	RM3+BM25 AND FINETUNED ROBERTA ROBERTA-BASE
STIS	STIS-E-7	RM3+BM25 AND ENCODER CONCAT GLOVE
STIS	STIS-E-8	RM3+BM25 AND ENCODER CONCAT GLOVE
STIS	STIS-E-9	RM3+BM25 AND ENCODER CONCAT GLOVE
STIS	STIS-E-10	RM3+BM25 AND FINETUNED BERT BERT-BASE-UNCASED
uhai	uhai-E-1	Query Fixing + L2R + Bert
uhai	uhai-E-2	Query Fixing + L2R
uhai	uhai-E-3	L2R + Bert
uhai	uhai-E-4	L2R
uhai	uhai-E-5	Query Fixing + bm25

Table 4: Evaluation results of the Japanese subtask runs.

	nDCG@3	nDCG@5	**nDCG@10**	nERR@3	nERR@5	nERR@10	Q-measure
KSU-J-5	0.388	0.403	0.448	0.283	0.448	0.477	0.498
KSU-J-1	0.362	0.381	0.421	0.295	0.423	0.453	0.473
ORGJ-J-3	0.407	0.413	0.421	0.325	0.450	0.470	0.484
uhai-J-10	0.403	0.406	0.415	0.312	0.447	0.466	0.484
ORGJ-J-2	0.402	0.405	0.415	0.328	0.447	0.467	0.483
ORGJ-J-6	0.379	0.386	0.406	0.321	0.423	0.447	0.464
ORGJ-J-1	0.382	0.396	0.405	0.308	0.426	0.452	0.464
ORGJ-J-7	0.380	0.386	0.401	0.323	0.430	0.452	0.471
ORGJ-J-4	0.365	0.377	0.400	0.318	0.409	0.433	0.452
uhai-J-9	0.369	0.382	0.393	0.301	0.417	0.441	0.461
uhai-J-6	0.369	0.375	0.389	0.293	0.418	0.439	0.455
ORGJ-J-5	0.362	0.363	0.377	0.288	0.415	0.434	0.452
ORGJ-J-8	0.357	0.363	0.373	0.289	0.404	0.425	0.437
uhai-J-7	0.350	0.352	0.368	0.272	0.410	0.431	0.453
uhai-J-8	0.346	0.350	0.362	0.278	0.392	0.414	0.432
KSU-J-3	0.114	0.117	0.119	0.045	0.136	0.145	0.151
KSU-J-7	0.114	0.117	0.119	0.045	0.136	0.145	0.151

Table 5: Evaluation results of the English subtask runs.

	nDCG@3	nDCG@5	**nDCG@10**	nERR@3	nERR@5	nERR@10	Q-measure
KSU-E-2	0.204	0.231	0.255	0.238	0.229	0.257	0.276
KSU-E-6	0.204	0.231	0.255	0.238	0.229	0.257	0.276
NIITableLinker-E-4	0.233	0.237	0.248	0.251	0.251	0.264	0.278
ORGE-E-2	0.219	0.225	0.238	0.240	0.235	0.250	0.264
uhai-E-5	0.219	0.225	0.238	0.240	0.235	0.250	0.264
NIITableLinker-E-10	0.221	0.226	0.237	0.238	0.235	0.248	0.264
STIS-E-2	0.230	0.228	0.237	0.217	0.248	0.255	0.264
ORGE-E-7	0.216	0.220	0.236	0.237	0.228	0.242	0.256
ORGE-E-8	0.224	0.230	0.233	0.238	0.244	0.255	0.264
NIITableLinker-E-1	0.201	0.211	0.231	0.228	0.221	0.239	0.257
NIITableLinker-E-5	0.214	0.227	0.230	0.234	0.230	0.247	0.258
uhai-E-3	0.209	0.214	0.227	0.237	0.223	0.234	0.249
STIS-E-10	0.208	0.209	0.221	0.208	0.234	0.242	0.253
STIS-E-1	0.201	0.201	0.221	0.199	0.227	0.234	0.249
uhai-E-1	0.200	0.209	0.219	0.232	0.213	0.225	0.239
NIITableLinker-E-2	0.202	0.205	0.219	0.235	0.217	0.230	0.244
ORGE-E-1	0.202	0.205	0.219	0.235	0.217	0.230	0.244
NIITableLinker-E-9	0.202	0.205	0.218	0.235	0.217	0.230	0.243
uhai-E-4	0.198	0.197	0.216	0.223	0.209	0.218	0.234
ORGE-E-4	0.192	0.201	0.213	0.226	0.207	0.224	0.238
ORGE-E-5	0.195	0.202	0.213	0.230	0.201	0.215	0.228
STIS-E-3	0.189	0.195	0.211	0.202	0.202	0.214	0.226
ORGE-E-6	0.171	0.191	0.205	0.221	0.189	0.212	0.226
uhai-E-2	0.173	0.178	0.203	0.213	0.184	0.194	0.213
NIITableLinker-E-3	0.192	0.194	0.203	0.219	0.209	0.217	0.230
STIS-E-6	0.165	0.175	0.197	0.187	0.182	0.194	0.211
NIITableLinker-E-6	0.157	0.168	0.193	0.212	0.157	0.171	0.191
STIS-E-4	0.172	0.171	0.192	0.185	0.190	0.199	0.212
NIITableLinker-E-7	0.173	0.180	0.190	0.185	0.189	0.205	0.219
NIITableLinker-E-8	0.171	0.176	0.190	0.204	0.180	0.193	0.206
ORGE-E-3	0.144	0.154	0.180	0.192	0.151	0.169	0.190
STIS-E-5	0.155	0.151	0.177	0.171	0.175	0.181	0.198
STIS-E-7	0.167	0.163	0.172	0.164	0.186	0.192	0.201
STIS-E-8	0.151	0.153	0.171	0.165	0.174	0.182	0.195
STIS-E-9	0.104	0.118	0.151	0.149	0.115	0.130	0.149
KSU-E-4	0.062	0.059	0.052	0.051	0.065	0.066	0.068
KSU-E-8	0.050	0.043	0.039	0.025	0.060	0.061	0.063