Overview of the VisEx task at NTCIR-9

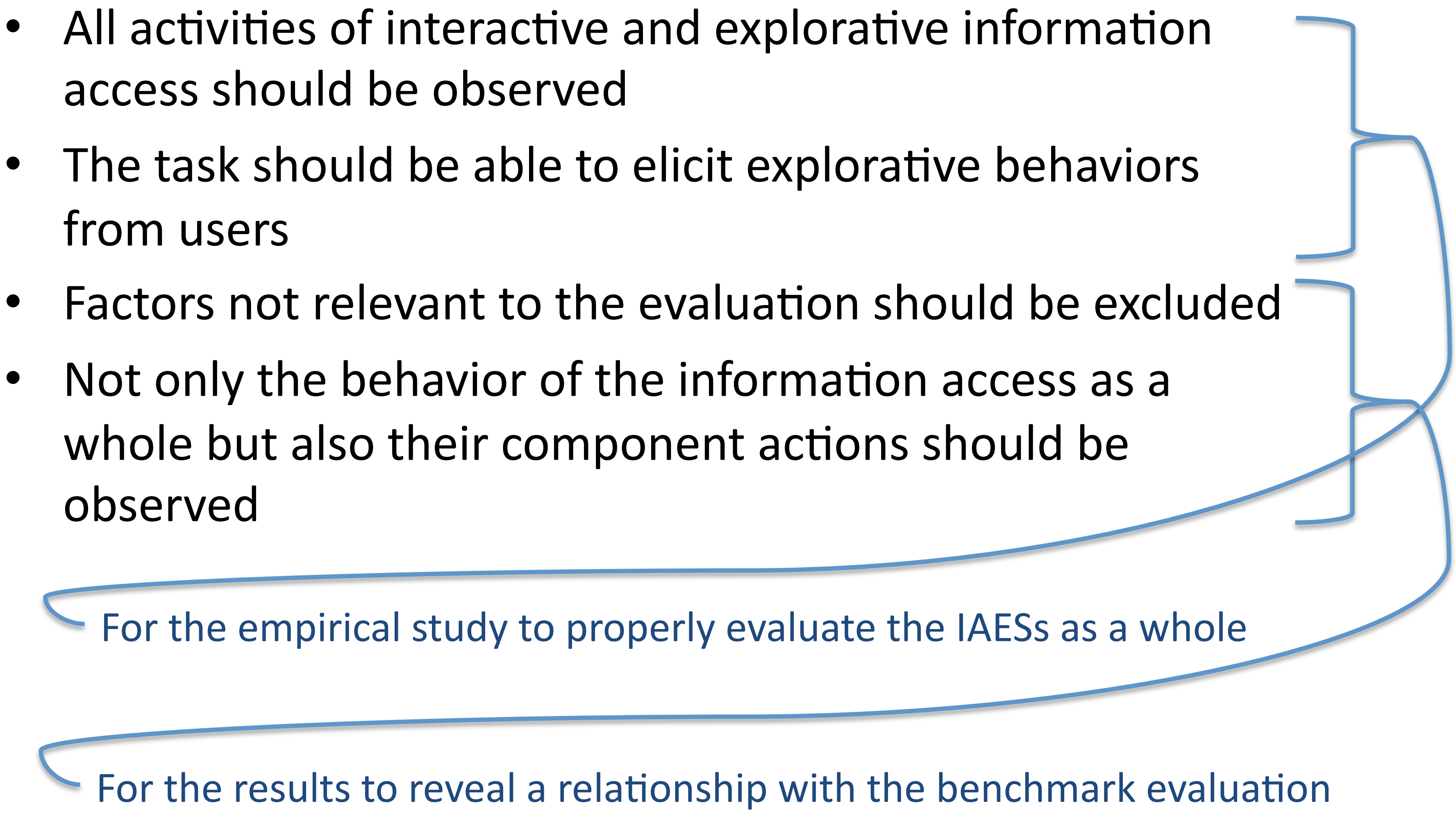
Tsuneaki Kato (The University of Tokyo)
Mitsunori Matsushita (Kansai University)
Hideo Joho (University of Tsukuba)

Objective

- To establish an efficient and effective framework for objective evaluation of interactive and explorative information access environment systems (IAESs)
  by bridging empirical user studies and benchmark tests

- To conduct sophisticated evaluation based on empirical user studies in order to acquire more useful and richer data, which is expected to reveal some relationship between two evaluation approaches

Policy



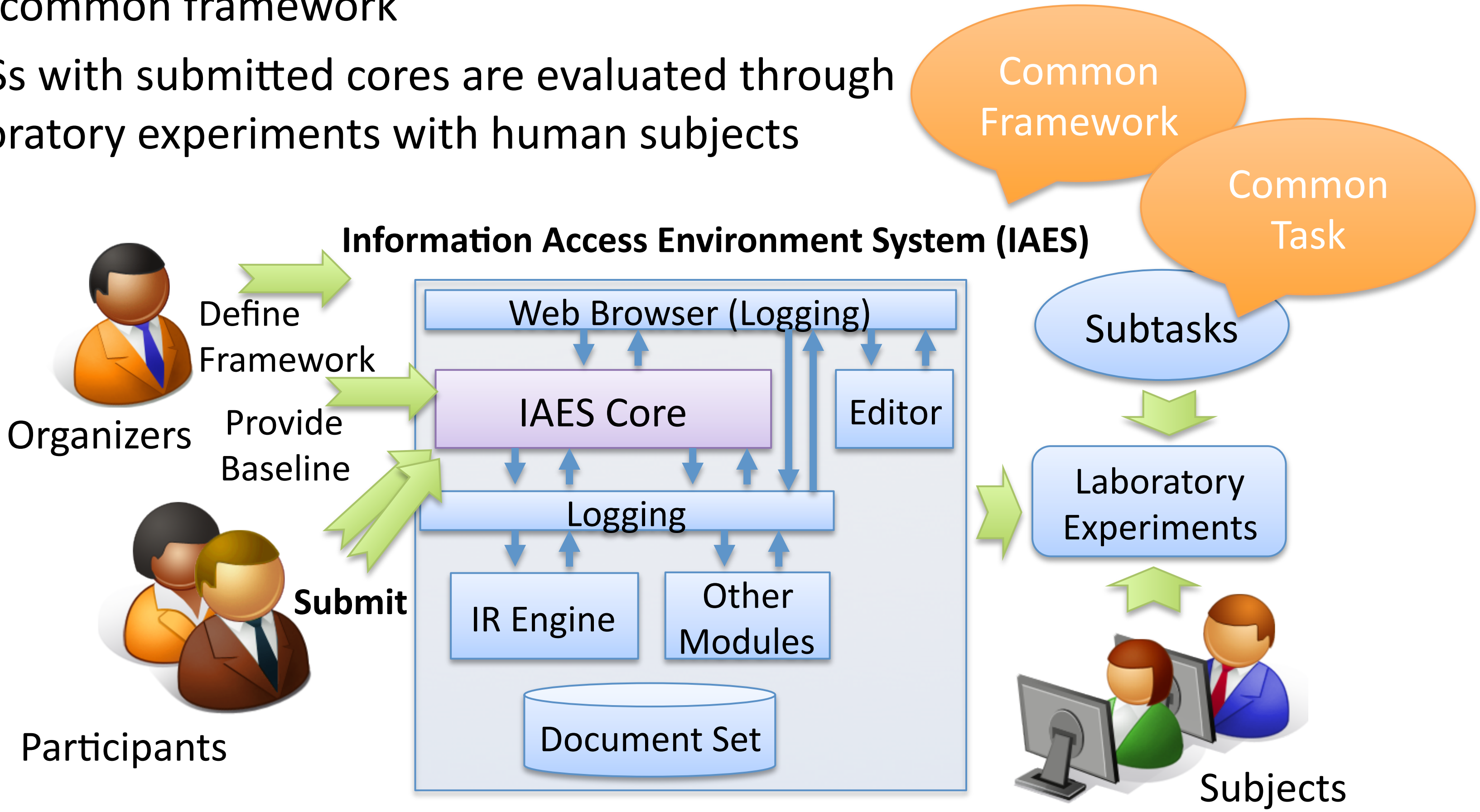
- All activities of interactive and explorative information access should be observed
- The task should be able to elicit explorative behaviors from users
- Factors not relevant to the evaluation should be excluded
- Not only the behavior of the information access as a whole but also their component actions should be observed

For the empirical study to properly evaluate the IAESs as a whole

For the results to reveal a relationship with the benchmark evaluation

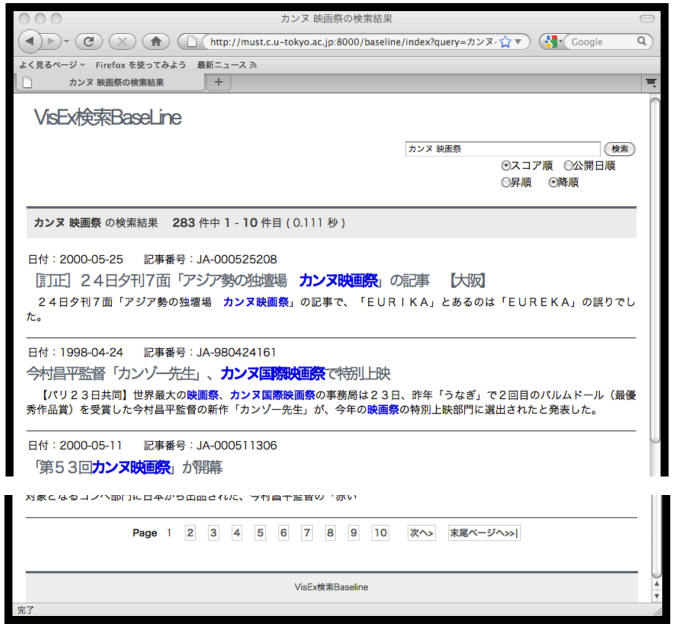
Framework

- A common framework is postulated for explorative information access environment systems (IAESs).
- The participants submit a core of an IAES, which works in the common framework
- IAESs with submitted cores are evaluated through laboratory experiments with human subjects



IAES Modules

- IR engine:
  Apache Solr full-text search server
- Web browser: Firefox browser
- Editor:
  developed as an add-on of Firefox
  - HTML document editor
  - Logger of users’ actions in the browser and the editor itself
- Baseline system
  - Provided by the organizers
  - Reference to be compared with submitted systems
  - Similar to ordinary web search engine
  - Allows users keyword-based retrieval



カンヌ 映画祭 の検索結果　283 件中 1 - 10 件目 (0.111 秒)

日付：2000-05-25　記事番号：JA-000525208

［訂正］２４日夕刊7面「アジア勢の独壇場　カンヌ映画祭」の記事　【大阪】

２４日夕刊７面「アジア勢の独壇場　カンヌ映画祭」の記事で、「ＥＵＲＩＫＡ」とあるのは「ＥＵＲＥＫＡ」の誤りでした。

日付：1998-04-24　記事番号：JA-980424161

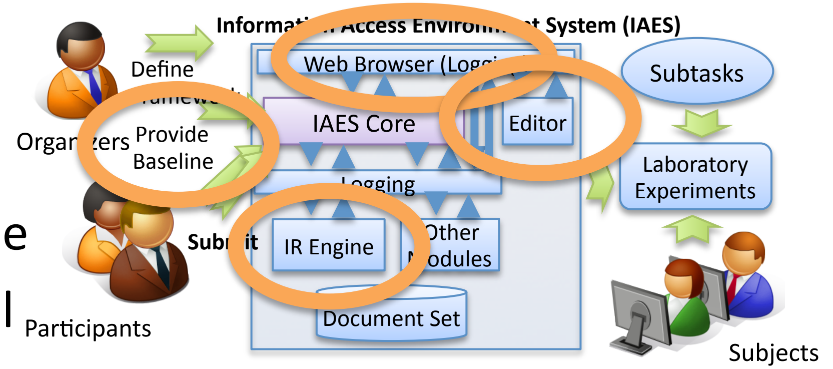
今村昌平監督「カンゾー先生」、カンヌ国際映画祭で特別上映

【パリ２３日共同】世界最大の映画祭、カンヌ国際映画祭の事務局は２３日、昨年「うなぎ」で２回目のパルムドール（最優秀作品賞）を受賞した今村昌平監督の新作「カンゾー先生」が、今年の映画祭の特別上映部門に選出されたと発表した。

日付：2000-05-11　記事番号：JA-000511306

「第５３回カンヌ映画祭」が開幕

対象となるコンペ部門に日本から出品された、今村昌平監督の「赤い

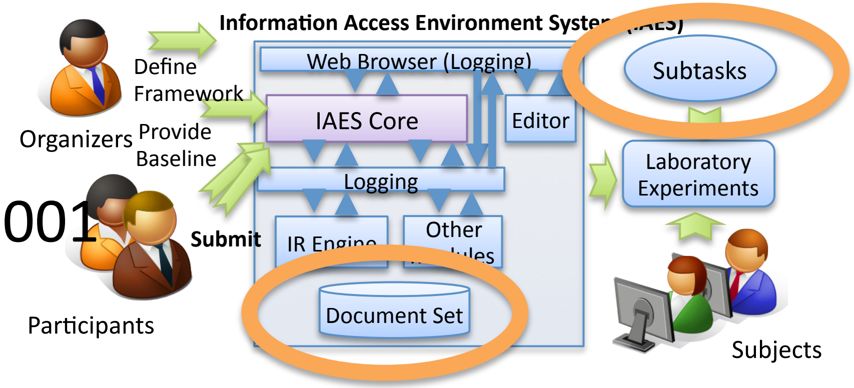


(Sub)Tasks and Document Set

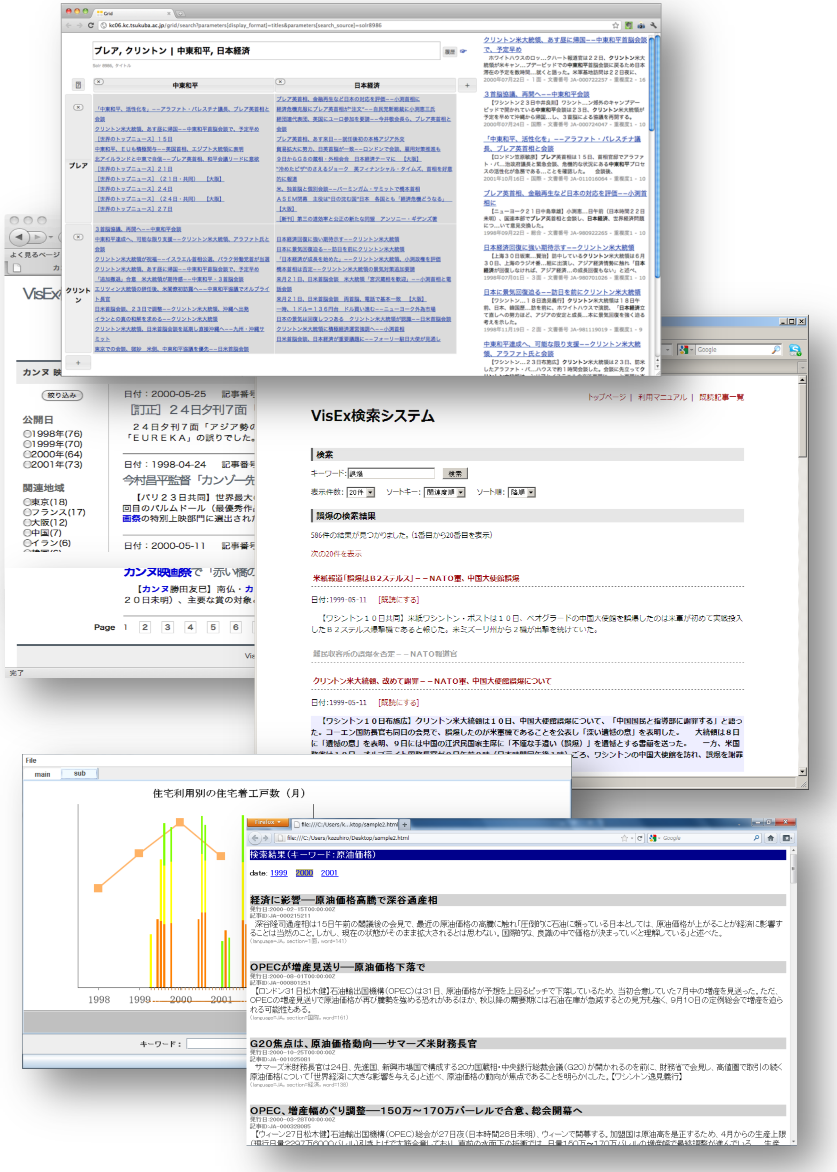
The task asks the subjects to collect information on a given topic and compile it into a report using a given IAES

- Event Collection Subtask
  - requests the subjects to make a report on events specified as a topic by collecting the characteristics of the events
    - Airplane crashes that have happened in Asia
    - Nuclear tests that have been conducted in different countries around the world
- Trend Summarization Subtask
  - requests the subjects to make a report summarizing trends related time-series statistical information given as a topic
    - The situation about gasoline
    - The evaluation of the Cabinet

Mainichi newspapers from 1998 to 2001 were used as the document set



Systems Submitted



ブレア, クリントン | 中東和平, 日本経済

VisEx検索システム

トップページ | 利用マニュアル | 既読記事一覧

検索

キーワード: 誤爆 検索

表示件数: 20件 ソートキー: 関連度順 ソート順: 降順

誤爆の検索結果

586件の結果が見つかりました。(1番目から20番目を表示)

次の20件を表示

米紙報道「誤爆はＢ２ステルス」－－ＮＡＴＯ軍、中国大使館誤爆

日付:1999-05-11 [既読にする]

【ワシントン１０日共同】米紙ワシントン・ポストは１０日、ベオグラードの中国大使館を誤爆したのは米軍が初めて実戦投入したＢ２ステルス爆撃機であると報じた。米ミズーリ州から２機が出撃を続けていた。

難民収容所の誤爆を否定－－ＮＡＴＯ報道官

クリントン米大統領、改めて謝罪－－ＮＡＴＯ軍、中国大使館誤爆について

日付:1999-05-11 [既読にする]

【ワシントン１０日布施広】クリントン米大統領は１０日、中国大使館誤爆について、「中国国民と指導部に謝罪する」と語った。コーエン国防長官も同日の会見で、誤爆したのが米軍機であることを公表し「深い遺憾の意」を表明した。 大統領は８日に「遺憾の意」を表明、９日には中国の江沢民国家主席に「不運な手違い（誤爆）」を遺憾とする書簡を送った。 一方、米国

住宅利用別の住宅着工戸数（月）

file:///C:/Users/kazuhiro/Desktop/sample2.html

検索結果(キーワード:原油価格)

date: 1999 2000 2001

経済に影響──原油価格高騰で深谷通産相

発行日:2000-02-15T00:00:00Z
記事ID:JA-000215211

深谷隆司通産相は15日午前の閣議後の会見で、最近の原油価格の高騰に触れ「圧倒的に石油に頼っている日本としては、原油価格が上がることが経済に影響することは当然のこと。しかし、現在の状態がそのまま拡大されるとは思わない。国際的な、良識の中で価格が決まっていくと理解している」と述べた。
(language=JA, section=1面, word=141)

OPECが増産見送り──原油価格下落で

発行日:2000-08-01T00:00:00Z
記事ID:JA-000801251

【ロンドン31日松木健】石油輸出国機構(OPEC)は31日、原油価格が予想を上回るピッチで下落しているため、当初合意していた7月中の増産を見送った。ただ、OPECの増産見送りで原油価格が再び騰勢を強める恐れがあるほか、秋以降の需要期には石油在庫が急減するとの見方も強く、9月10日の定例総会で増産を迫られる可能性もある。
(language=JA, section=国際, word=161)

G20焦点は、原油価格動向──サマーズ米財務長官

発行日:2000-10-25T00:00:00Z
記事ID:JA-001025081

サマーズ米財務長官は24日、先進国、新興市場国で構成する20カ国蔵相・中央銀行総裁会議(G20)が開かれるのを前に、財務省で会見し、高値圏で取引の続く原油価格について「世界経済に大きな影響を与える」と述べ、原油価格の動向が焦点であることを明らかにした。【ワシントン逸見義行】
(language=JA, section=経済, word=138)

OPEC、増産幅めぐり調整──150万～170万バーレルで合意、総会開幕へ

発行日:2000-03-28T00:00:00Z
記事ID:JA-000328085

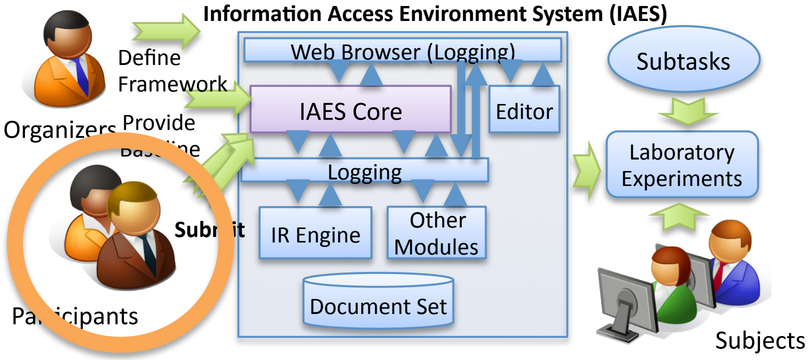
【ウィーン27日松木健】石油輸出国機構(OPEC)総会が27日夜(日本時間28日未明)、ウィーンで開幕する。加盟国は原油高を是正するため、4月からの生産上限

Four teams participated
Each team submitted one system

- University of Tsukuba (Grid)
- Kansai University (KN)
- Tokyo Metropolitan University (TM2011)
- The University of Tokyo (UTLIS)

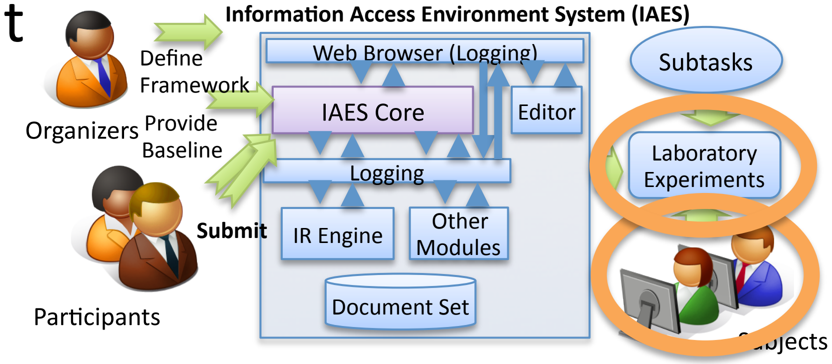
Two types of the submissions

- Proposals of a novel interface
- Experiment on the effects of some function



Experiment Design

- One unit of experiment is a combination of a system and a subtask
  - Eight units of the experiment were conducted including two units of the baseline system
- Five subjects attend each unit
  - Each subject attends just one unit
    … between-subjects experiment design
- Each subtask consists of four topics and one training topic
- Pre-experiment, post-session (Five times), and post-experiment questionnaire surveys were conducted



Data Obtained

Data obtained through the experiments include

- Reports made by the subjects
  - Main products of the information access activities
- Log records
  - Dynamic process of the activities
- Data collected using questionnaire surveys
  - Subjective impressions of the subjects involved

It is expected to be possible to comprehensively grasp the complex activities of interactive information access, and to understand the roles of IAESs in those activities by analyzing the data individually and synthetically

Outcome

- Extensive range of data was obtained on users' behavior and their impression
  - We are still on the way of the analysis
  - The basic framework was confirmed to be promising
- Every team obtained valuable data for the evaluation of the submitted system
- It was more difficult than expected to understand the obtained data and draw a clear picture from it
  - Especially from the viewpoint of establishing an evaluation framework

Possible reasons for the difficulty

- Great diversity of behavior among the subjects and too few subjects
  - The characteristics of IAESs were buried in such diversity
- Difficulty to quantify the suitability of the products and processes of interactive information access
  - No clear criterion for what is the ideal
- Some lack of sophistication on the mechanisms for capturing log records

Example of Data Analysis (Report Analysis)

Number of articles referred to in reports

		1			2			3			4		
		min	max	med	min	max	med	min	max	med	min	max	med
Event	Baseline	4	12	8	3	10	6	9	14	11	9	18	12
	Grid	5	11	6	4	13	8	6	15	8	7	14	12
	TM2011	5	15	10	4	12	9	8	18	13	11	18	14
	UTLIS	4	12	10	7	12	8	6	11	9	8	13	9
Trend	Baseline	4	8	6	5	12	9	5	14	7	5	9	8
	Grid	4	8	6	3	16	7	4	13	8	3	13	12
	KN	4	10	10	2	23	10	8	22	9	9	13	10
	UTLIS	5	10	6	4	9	9	6	12	8	6	12	8

Similarity to the baseline in terms of retrieved articles

		1	2	3	4
EVENT	Grid	0.48	0.28	0.50	0.57
	TM2011	0.44	0.41	0.70	0.75
	UTLIS	0.54	0.28	0.67	0.67
TREND	Grid	0.76	0.60	0.20	0.70
	KN	0.50	0.35	0.41	0.78
	UTLIS	0.72	0.66	0.38	0.79

Example of Data Analysis (Log Analysis)

The proportion of knowledge compilation time

		1	2	3	4
EVENT	Baseline	0.42	0.26	0.38	0.31
	Grid	0.45	0.38	0.44	0.43
	TM2011	0.53	0.32	0.40	0.43
	UTLIS	0.44	0.33	0.31	0.36
TREND	Baseline	0.55	0.60	0.49	0.41
	Grid	0.58	0.56	0.55	0.53
	KN	0.41	0.36	0.31	0.34
	UTLIS	0.56	0.58	0.58	0.41

Knowledge compiling time:
the time when the editor tab is active
an approximation of the time taken by the user
to compile information and write it into a report

Example of Data Analysis (Questionnaire)

System Evaluation through Questionnaire Surveys

		Usability	Functionality	Efficiency
EVENT	Baseline	5.8	4.0	4.0
	Grid	2.8	3.8	3.2
	TM2011	5.6	3.8	4.2
	UTLIS	5.8	4.2	5.2
TREND	Baseline	4.0	3.4	3.2
	Grid	5.8	5.0	5.2
	KN	5.4	4.4	4.6
	UTLIS	5.0	4.o	4.8

The average values of the subjects' answers to questions, each of which used a seven-point Likert scale with a score of seven meaning the best evaluation

Conclusion

- The VisEx task was conducted, which aimed to establish a framework for evaluating interactive and explorative information access environments
- Four teams participated
- The basic framework was confirmed to be promising
  - Extensive range of data was obtained on several aspects of complex behaviors of interactive information access
- A lots of lessons have been learned
  - The task should be more difficult in order to derive explorative behaviors of users
  - The diversity of user behavior should be reduced somehow
  - More sophisticated log-taking mechanism or principle is expected