

統合型eテスト・システム の開発と実践

ソムアン ポクポン, 植野 真臣
電気通信大学大学院情報システム学研究所

eテストの利点

1. コンピュータによるテスト構成支援機能は、信頼性と妥当性の高いテスト構成を可能にする。
2. 適応型テストなど受検者の能力を推定しながらその能力に最適な項目を選択出題するような動的テストを実現し、テスト時間を短縮し、測定精度も向上させることができる。
3. 遠隔地にいる受検者の評価を行うことができる。
4. 遠隔地に分散するテスト構成者が、共同でアイテム・バンクを構築し、テスト構成を行うことができる。
5. アイテム・バンクの領域からランダムに項目の抽出が行え、テスト出題バイアスを減少できる。
6. マルチメディア技術を用いることにより、動画や音声を用いたテスト項目を作成でき、これまで測定することのできなかつた能力を測定できるテスト項目を構築、実施できる。
7. 回答所要時間や回答変容回数、回答見直し回数などペーパー・テストでは測定することができなかつたデータが測定できるようになり、テスト項目評価の多様性をもたらす、受検者評価に補助的に用いることにより測定精度を向上させることができる。
8. テスト実施に際して、テストの配布、回答の回収が自動的に行える。
9. テスト採点が、自動化もしくは半自動化される。即時的にテスト結果をフィードバックできる。
10. 障害を持つ受検者への対策が容易に行える。例えば、字を大きくしたり、ヘッドホンをつけさせて音を大きくしたり、特別な入力装置(手以外で入力できる入力装置)、出力装置(点字など)を用いることにより障害者の受検を可能にすることができる。
11. 大量のペーパー・テストの束を管理することに比較し、電子ファイル上のテストは管理が容易でセキュリティも高くできる。
12. ペーパー・テストに比較し、テストの印刷、運搬のコストが減少する。

表1. eテストニング研究の特徴の比較

利点	1	2	3	4	5	6	7	8	9	10	11	12
永岡 (2000)		×	×	×		×						
Conejo et al. (2004)	×			×	×	×						
He and Tymms (2005)	×			×	×		×					
Kerejeta et al. (2005)	×	×		×		×	×					
Ueno (2005)		×				×	×					
仲林ほか (2006)	×	×		×	×	×						
本システム												

:有り × :なし :どちらともいえない



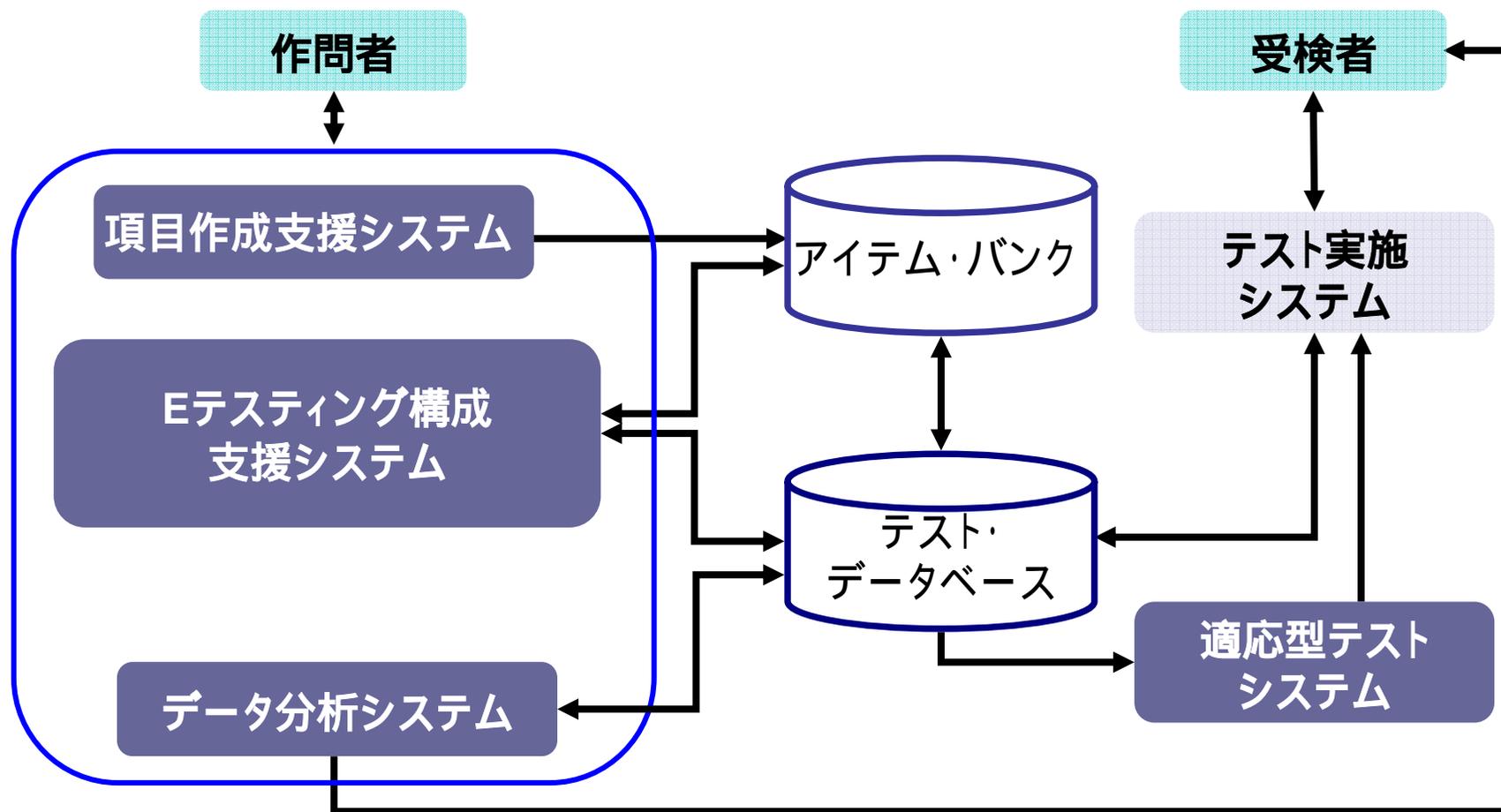
研究目的

本研究では、eテストの利点をすべて統合的に有し、実用的なeテスト・システムの開発を目的とする。

統合的システムを開発することの利点として以下の点が挙げられる。

1. 過去に蓄積されたテスト・データが、各機能に一貫して自動的に配分され、テスト分析、項目分析、テスト構成支援、適応型テストなど様々な機能に利用することができる。
2. 多様な機能を持ち、一つのシステムで多様な目的(入試、能力測定、形成的評価、自己評価やeラーニング)のテストを実施することができる。

統合的eテストイング・システム



項目作成支援システム

■ マルチメディア項目

- テキスト
- 画像
- 動画
- 音声





eテストイング構成支援システム

本システムにおける技術的に最も特徴的な機能は、eテストイング構成支援システムである。

eテストイング構成支援システムの特徴

1. 遠隔地にいる複数のテスト構成者の協働テスト構成を実現している。
2. 構成されているテストの得点分布およびテスト所要時間を予測・提示する機能を持つ。

eテストニング構成支援システム

Test Name: test
 Test ID: 078
 Subjects: 1
 Definition: ttt
 Expect Required Time: now showing in predictive response time
 120 seconds

Author: test2
 Invited Co-Authors: Polpong

number of items in the test = 13
 number of test-author(s) = 2

Item ID	Number of selected trial	Average of required time	Rate of correct answers
06006MPA100	21	00:00:30	0.8142857142857143
06006MT063	40	00:00:15	0.673913043478261
06006MT068	20	00:00:16	0.632183909045977
06006MT062			

Todays Date: 10/17/2007 11:50:46

agent agent recommends 06006MT068
 test2 item please agent test efficiency is improved after delete 06006MPA100
 agent test efficiency is improved after delete 06006MT037
 agent test efficiency is improved after add 06006MT068
 agent agent recommends 06006MT068
 test2 item please agent test efficiency is decreased after delete 06006MT068
 agent test efficiency is improved after add 06006MT049
 agent agent recommends 06006MT049
 test2 item please agent test efficiency

06006MT063 00:00:25
 06006MT068 00:00:13
 06006MT069 00:00:09
 06006MT072 00:00:20
 06006MT074 00:00:30
 06006MT077 00:00:10
 06006MT087 00:00:08

Predictive Response Time Distribution
 Predictive Score Distribution

得点と所要時間の予測モデル選定

得点分布

1. 混合ベータ二項分布
2. 項目反応理論を用いたテスト得点分布
 - ラッシュモデル
 - 2パラメータ・ロジスティック・モデル
3. 二項分布 (Ueno 2005)
4. 混合二項分布 (Ueno 2005)
5. ベータ二項分布 (Bradley and Robert 1990)
6. 切断指数分布 (永岡 2000)

所要時間

1. 正規分布
2. 対数正規分布 (Tissen 1983)
3. 拡張ガンマ分布 (Ueno and Nagaoka 2005)
4. ワイブル分布 (Roskam 1997)

モデルの採択結果

■ 予測得点分布モデルの選択

- 項目反応理論(2パラメータ・ロジスティック・モデル)を用いたテスト得点分布(ソナムアン・植野 2008)

$$p(x) = \int_{-\infty}^{\infty} \sum_{i=1}^m \frac{1}{m} \binom{m}{x} p(u_i | \theta_j) p(\theta_j) d\theta_j$$

■ 予測所要時間分布モデルの選択

- 拡張ガンマ分布(Ueno and Nagaoka, 2005)

$$f(t) = \frac{t^{\alpha-1} \exp(-t/\beta)}{\beta^{\alpha} (\alpha-1)!}$$

テスト実施システム

- テスト配信システムとも呼ばれ、指定されたIDの項目をeテストング・サーバーから各受検者の端末にインターネットを通じて配信するとともに、受検者からの回答や反応履歴を獲得し、サーバーに格納する機能を持つ。

Show Item - Microsoft Internet Explorer

Multiple choice

Video
wmv



Command :

Question :
上のビデオを見て、アメリカ人家のテーブルはどこからもらいましたか答えなさい。ただし、もっとも適切な選択肢を下より一つえらびなさい。

Answer

- アメリカで買った
- 日本で買った
- 自分で作った Correct!!!!
- 男の人からもらった

データ分析システム

各受検者のテスト・データの解析

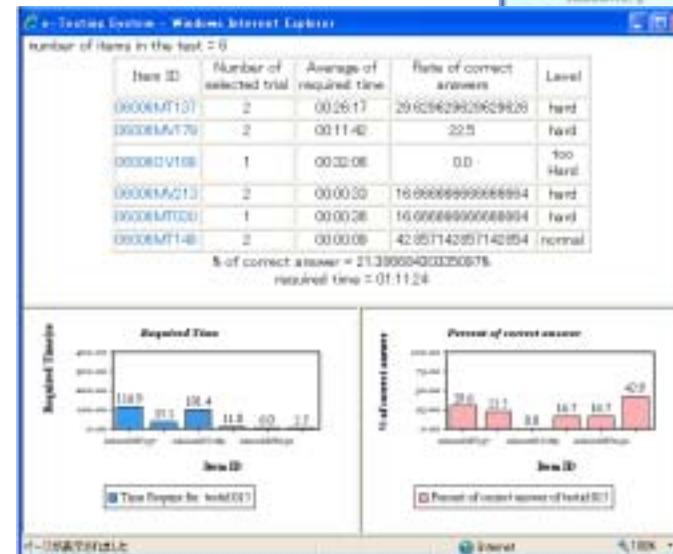
- 正誤,
- テスト得点,
- 項目所要時間,
- 回答書き直し履歴,
- 見直し履歴,
- 平均テスト得点,
- 平均項目所要時間

テストについての解析

- 各項目ID,
- 正答率,
- 項目平均所要時間,
- 項目平均回答書き直し回数,
- テスト得点分布,
- テスト所要時間分布

Test Name	Required Time	Historical Log Date	Test times	Date
Test	time required	Show	1	2008-02-14
Test	time required	Show	2	2008-02-14

Item ID	Correct Answer	Student Answer	Change Frequency	Mark	Required Time
0000MT101	0113	0113	0	0	000000
0000MT102	0113	0113	0	0	000000
0000MT103	0113	0113	0	0	000000
0000MT104	0113	0113	0	0	000000
0000MT105	011224	011224	1	0	000000
0000MT106	011234	011234	1	0	000000
0000MT107	0113	0113	0	1	000000





データ分析システム

項目反応理論のパラメータ推定法

1. 解析的ベイズ推定(Ueno, 2004), (高橋・植野, 2006),
2. Bayesian Parameter Estimation (Mislevy, 1986),
3. Marginal Maximum Likelihood (Bock& Aitkin, 1981),
4. Gibbs Sampler (Albert, 1992)



適応型テストシステム

様々な項目選択基準を選択して用いることができるようになってきている。

- フィッシャー情報量最大化基準(Lord & Novick, 1968)
- 能力推定値-項目難易度距離の最小化基準(Owen, 1975)
- カルバックライブラー情報量最小化基準(Chang & Ying, 1996)
- 十分ベイズ基準(van der Linden, 1998)



実践と評価

本研究は単に理論研究だけでなく、実際に役に立つシステム開発を行ってきた。

20人の大学教員に実際に1年間使用してもらい、その後、アンケート調査を行うことによって、本システムが単なるプロトタイプではなく、実用的なシステムであることを示した。

表2. 本システムの使用状況

授業名	形式	用途	期間	真偽式項目	多肢選択式項目	単答式項目	論述式項目	受験者	受検者数
基礎統計	講義	演習・期末	後期	0	116	11	0	B1	10
社会知能情報学基礎	講義	演習	前期	7	68	2	5	B3	60
教員育成論	講義	演習	通年	0	0	0	15	B3-4	42
知識創産システム論	講義	演習	後期	9	17	0	4	M1-2	18
安全規制とマネジメント	講義	演習	通年	5	0	0	12	M1-2	27
工学基礎1	実験授業	演習	前期	0	36	0	0	B1	13
工学基礎2	実験授業	演習	後期	0	56	0	0	B1	12
e-Bayesian Statistics	eラーニング	演習	前期	0	20	0	0	M1-2	20
e-ベイズ統計	eラーニング	演習・期末	通年	0	164	28	4	M1-2	30
e-情報科教育法	eラーニング	演習	前期	4	201	0	4	B3-4	32
e-情報技術基礎	eラーニング	演習	前期	0	2	0	0	B3-4	23
e-情報社会と情報倫理 A	eラーニング	演習・期末	通年	47	298	1	16	B3-4	95
e-情報社会と情報倫理B	eラーニング	演習・期末	通年	37	216	6	3	B3-4	96
e-統計工学基礎	eラーニング	演習・期末	前期	0	116	11	0	B3-4	82
ナレッジマネジメント論	eラーニング	演習	後期	20	10	2	14	M1-2	17
e-技術と社会：技術者倫理入門	eラーニング	演習	前期	0	10	0	0	B3-4	96
e-経営情報学基礎	eラーニング	演習	前期	0	90	0	0	B3-4	48
e-情報技術基礎 2	eラーニング	演習	前期	0	186	0	0	B3-4	42
視聴覚メディア論	遠隔授業	演習・期末	前期	7	68	2	5	B3	31
International Industry Politics	国際遠隔授業	評価テスト	5時間	0	6	0	0	B4	20

システムの使用状況

- 講義形式の授業での評価
- eラーニングでの評価
- 遠隔授業での評価
- 国際遠隔授業での評価





アンケート調査

マルチメディア項目作成機能を用い、有効と感じましたか。

テスト構成支援システムは有効でしたか。

テスト得点の予測機能は使いやすかったですか。

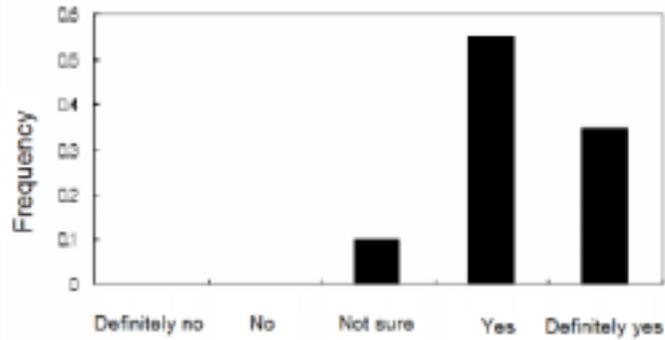
テスト所要時間の予測機能は使いやすかったですか。

データ分析機能による項目分析、テスト分析は使いやすかったですか。

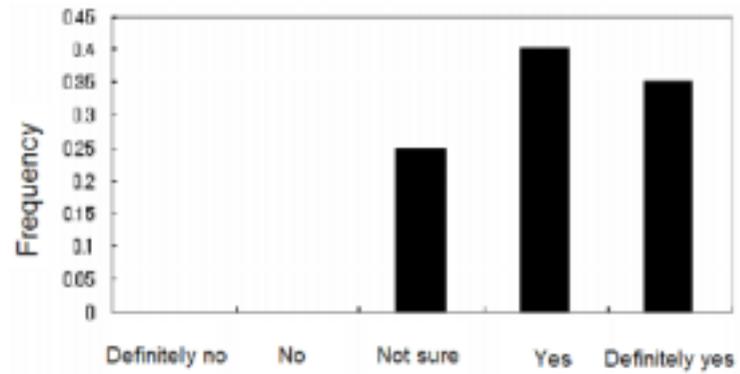
データ分析機能による受検者分析は使いやすかったですか。

アンケートの結果

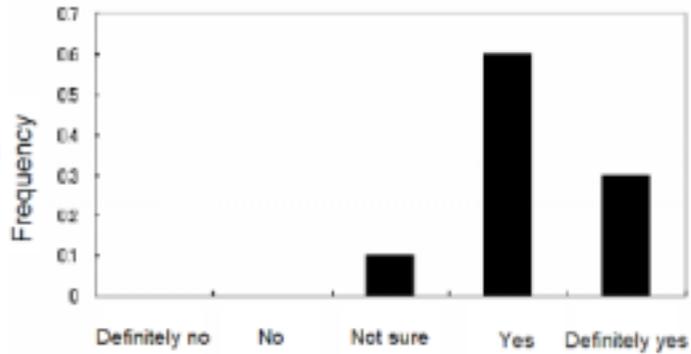
1.



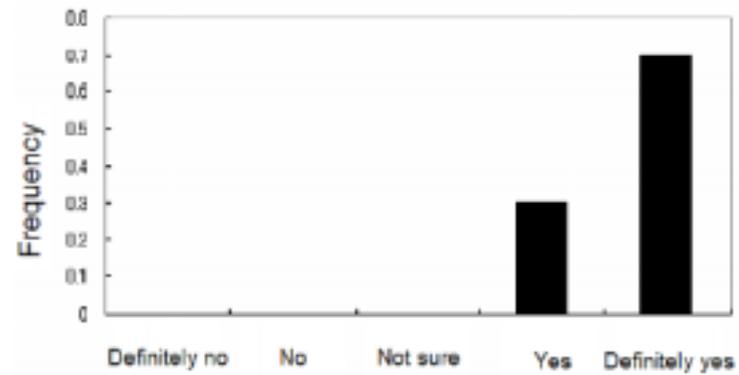
4.



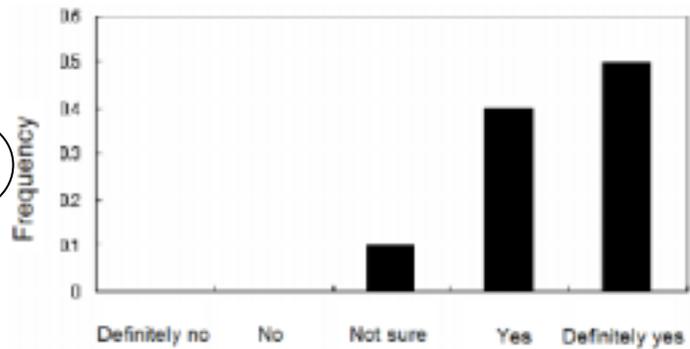
2.



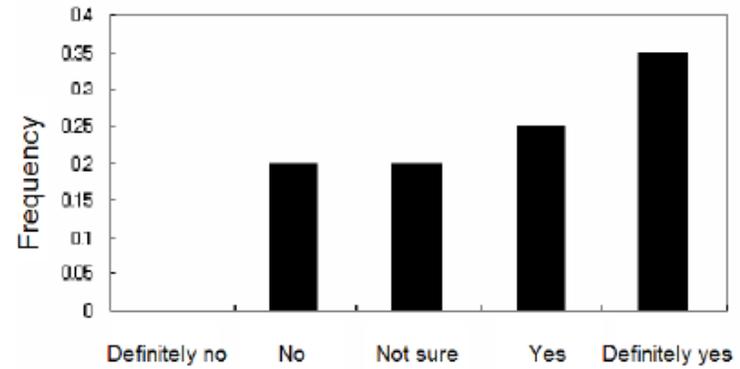
5.



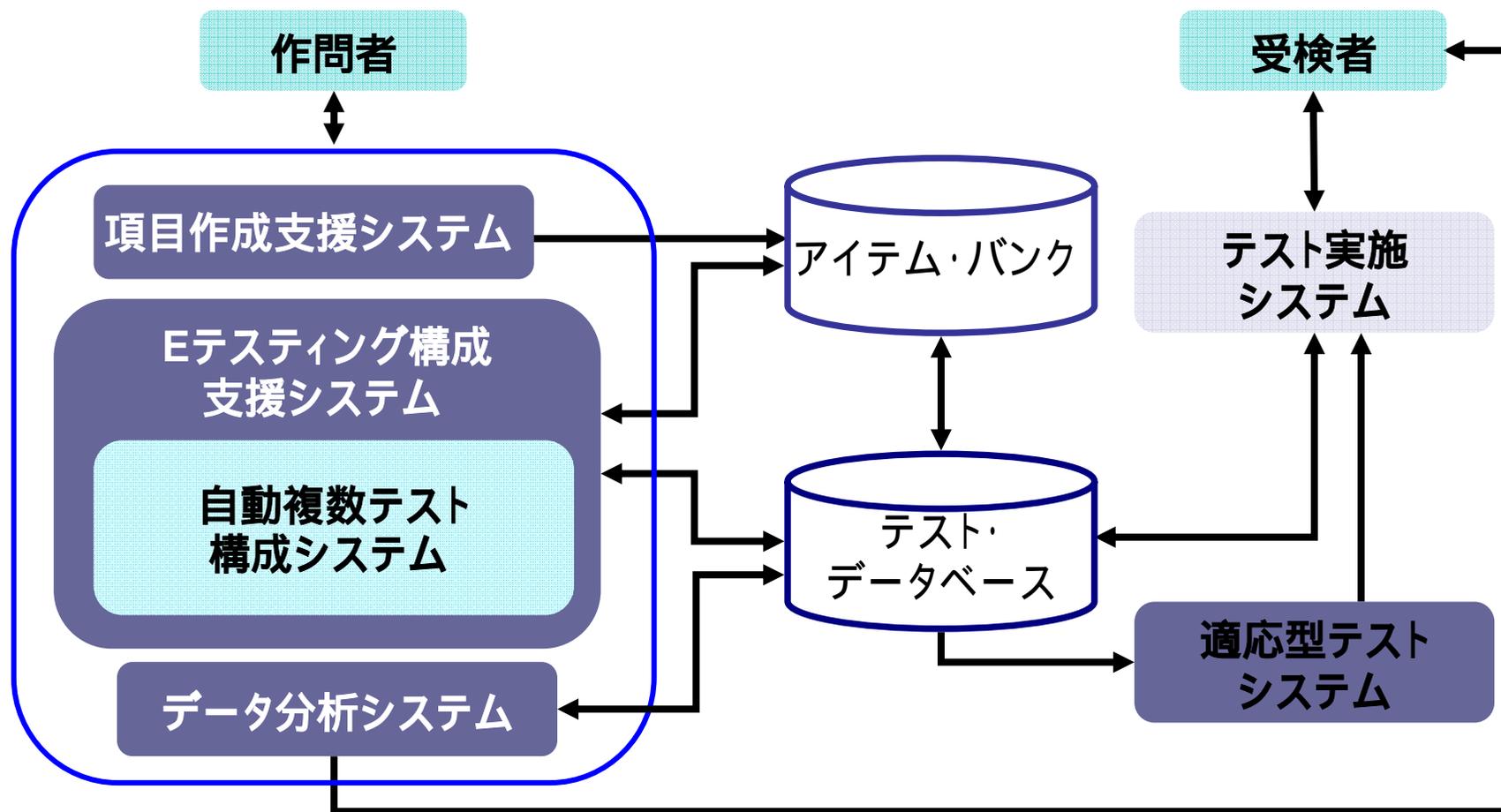
3.



6.



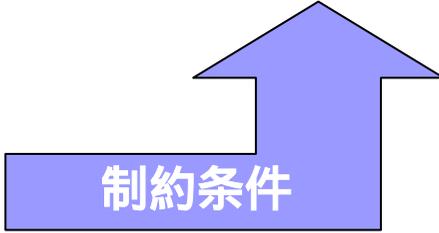
統合的eテストイング・システム



自動複数テスト構成システム



The screenshot shows the user interface of the system. On the left, there are various settings and filters. The main part of the screen is a table with columns for '科目' (Subject), '項目ID' (Item ID), '項目名' (Item Name), '項目数' (Item Count), and '項目内容' (Item Content). The table lists various items for different subjects like '国語' (Japanese), '算数' (Mathematics), and '理科' (Science).



制約条件

- 例えば:
- 平均得点 60点
 - 所要時間 30分
 - 科目Aの項目数 10項目
 - etc....
- の条件

自動複数テスト構成システム

テスト構成支援システム

ファイル(F) 編集(E) 表示(V) 履歴(H) ブックマーク(B) ウィンドウ(W) ヘルプ(H)

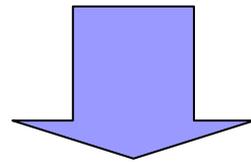
作成日	2008/06/20	
Ver.		
構成方法	マニュアル	
キーワード		
備考		
問題セット数		
項目実施回数 (下限値)		
テスト間類似度 (下限値)		
実質領域数 (下限値)	分野	低 <input type="text"/> ~ 高 <input type="text"/>
	大分類	低 <input type="text"/> ~ 高 <input type="text"/>
	中分類	低 <input type="text"/> ~ 高 <input type="text"/>
パラメータ (下限値)		
テスト情報量 (下限値)		
含否①		
含否②テスト情報量 (下限値)		
正答率	平均	低 <input type="text"/> ~ 高 <input type="text"/>
	標準偏差	低 <input type="text"/> ~ 高 <input type="text"/>
所要時間 (秒)	平均	低 <input type="text"/> ~ 高 <input type="text"/>
	標準偏差	低 <input type="text"/> ~ 高 <input type="text"/>
合計問題数		
実行		

タイプ	分野	分野別抽出数 (下限値)	大分類	中分類	抽出数 (下限値)		
小問	ストラテジ系	<input type="checkbox"/>	1 企業と法務	1 企業活動	<input type="checkbox"/>		
				2 法務	<input type="checkbox"/>		
			2 経営戦略	3 経営戦略マネジメント	<input type="checkbox"/>		
				4 技術戦略マネジメント	<input type="checkbox"/>		
				5 ビジネスインダストリ	<input type="checkbox"/>		
			3 システム戦略	6 システム戦略	7 システム企画	<input type="checkbox"/>	
					8 システム開発技術	<input type="checkbox"/>	
	4 開発技術	<input type="checkbox"/>	9 ソフトウェア開発管理技術	10 プロジェクトマネジメント	<input type="checkbox"/>		
				11 サービスマネジメント	<input type="checkbox"/>		
			6 サービスマネジメント	12 システム監査	<input type="checkbox"/>		
				13 基礎理論	<input type="checkbox"/>		
			7 基礎理論	<input type="checkbox"/>	14 アルゴリズムとプログラミング	15 コンピュータ構成要素	<input type="checkbox"/>
	16 システム構成要素	<input type="checkbox"/>					
	8 コンピュータシステム	17 システム構成要素			18 ハードウェア	<input type="checkbox"/>	
					19 ヒューマンインタフェース	<input type="checkbox"/>	
	9 技術要素	<input type="checkbox"/>			20 マルチメディア	21 データベース	<input type="checkbox"/>
						22 ネットワーク	<input type="checkbox"/>
					23 セキュリティ	<input type="checkbox"/>	
	中問	ストラテジ系			<input type="checkbox"/>		
		マネジメント系	<input type="checkbox"/>				
		テクノロジー系	<input type="checkbox"/>				



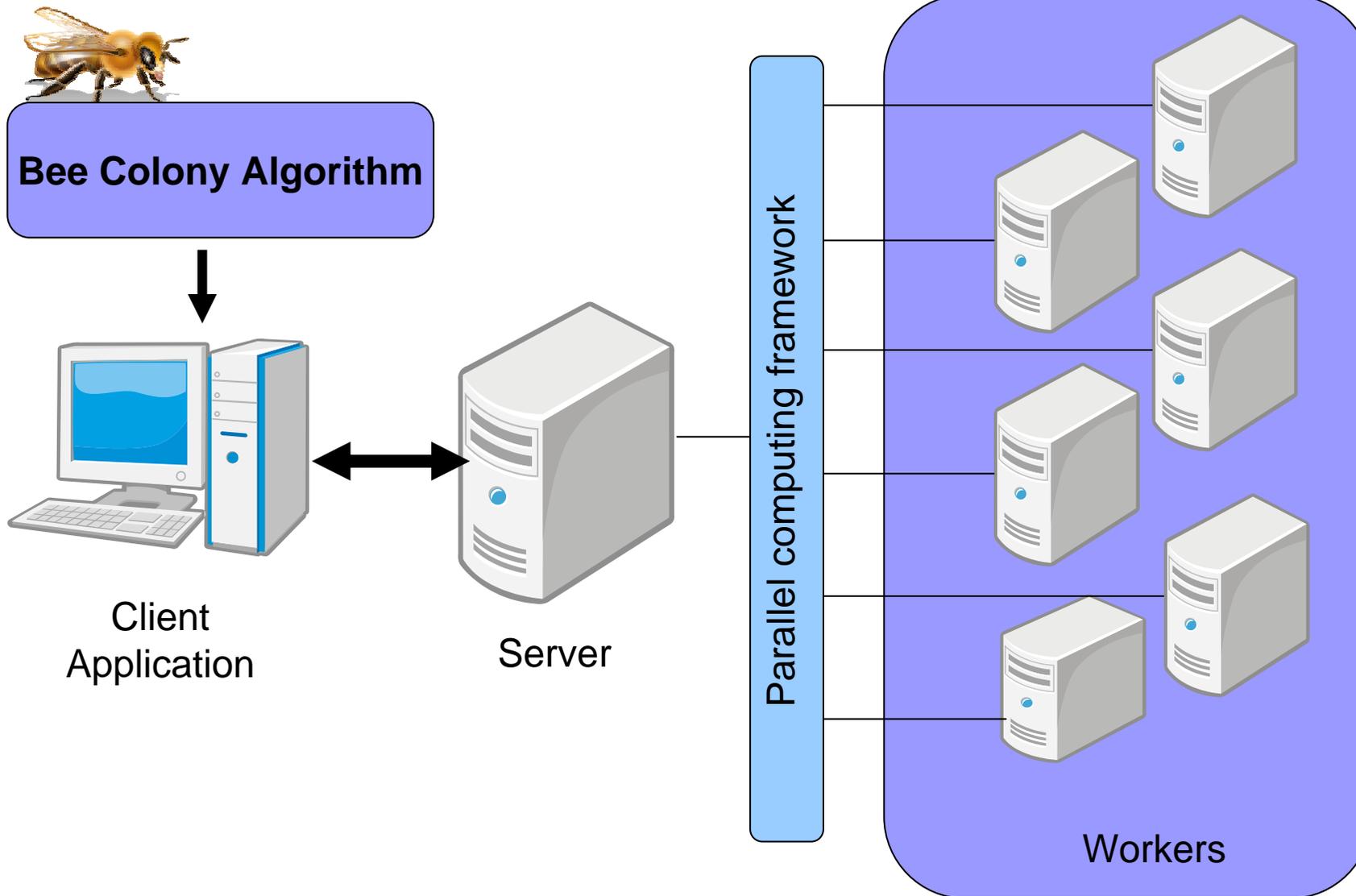
従来手法

- 従来の複数テスト構成アルゴリズム
 - 同時テスト構成 (Adema, 1990)
 - Big-Shadow Test (Linden, 2005)
 - 最大クリーク抽出法 (石井・植野, 2008)



- 複数の等質テストを構成するために時間がかかり過ぎる.

提案手法

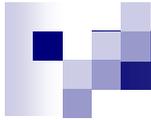




実験結果

アイテム・バンク1000項目, 40条件, テスト30項目

手法	CPU cores	テスト数	平均テスト情報量	平均テスト情報量の標準偏差	計算時間
従来手法	1	6	6.001	2.40×10^{-3}	> 3日間
提案手法	1	6	5.942	0.225	00:56:21
	2	6	5.956	0.303	00:47:73
	3	6	5.991	0.230	00:30:26
	4	6	5.821	0.248	00:25:08
	5	6	6.047	0.271	00:23:47

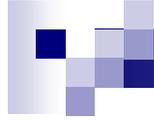


e テスティング

植野真臣・永岡慶三 共編



培風館



ご清聴ありがとうございました。