学生実験支援システムの構築と運用
－成績とシステムアクセス頻度、習熟の過程－

丸山 広 †　　中村 太一 ††

情報処理の技術はコンピュータの基礎技術から応用技術に進展している。その技術の進展の動向に応えるためにテキストマイニングを使い Web の情報を分析する実験カリキュラムを開発した。この実験カリキュラムの運用に必要な、実験レポートの受領、実験に関する通知、実験に必要なデータやモジュールの配布およびテキストマイニングを実行環境で構成される支援システムを開発した。本稿では開発したシステムの概要、学生の成績と当該システムへのアクセス回数および Web ベースの支援システムに対する学生の習熟の変化について報告し、得られた知見について述べる。

Development and Operation of the IT assistance system for the program of experiment, -The relationship between the grade point of students and the frequency of access to the system, and a learning process ?

Hiroshi MARUYAMA‡　　Taichi NAKAMURA‡‡

Technologies of the information processing have been developing from a basic field to an application field. In order to support this technology development trend, we have constructed the program of experiment using a text mining technology to analyze the information in the Web. We have developed the IT assistance system which has several functions such as, a submission of a report from a student, a notification of the information for the experiment program, delivering the data and modules used in this course and a text mining execution environment. This paper describes the outline of the system, the relationship between the grade point of students and the frequency of access to the system, and a process of the students learned on the Web based assistant system.

1　まえがき

情報処理技術の発展と普及に伴い、情報処理の技術分野が基礎から応用へと広がっていることは衆知のことと考える。このような産業界の動向に対応するため、情報工学分野のカリキュラムでは、どのようにしてコンピュータが動作するのか、と言った原理や基礎技術の習得に加え、コンピュータを如何に利用し、私達の生活に資する付加価値を生み出すか、と言った利用技術の習得の重要性が認識されるようになっている。勿論、基礎技術は利用技術を学ぶために必須であり、その重要性は何なら変わるものではない。

利用技術を教えるにあたり、基礎技術の学習方法との違いを留意しておく必要があると考える。基礎技術は論理的あるいは物理的な事象に基づく知識体系で構成されているので、問題に対する答えは

† 東京工科大学工学部情報工学科中村太一研究室
†† 東京工科大学ＣＳ学部
‡Taichi-Nakamura Laboratory, The department of faculty-of-technology information engineering, Tokyo University of Technology
‡‡The School of Computer Science, Tokyo University of Technology

唯一であることが殆どである。他方、利用技術においては事象が確率の要素を含んでおり、得られた解釈の多義性のため結果の多様性を許容する必要があると考える。

換言すると、利用技術のカリキュラムでは、それぞれの学生が導き出した回答に、個別に解釈を与える個別指導が求められるのではないかと考える。加えて、コンピュータ操作のスキルや講義内容に対する習熟度のばらつきが大きいことも個別指導が求められる要因である。

しかしながら、現実的には教員スタッフの数と学生数の関係において運用の稼動を考えると、当大学での教育はマスプロ教育にならざるを得ない。このような状況において学生に個別指導が求められるコンピュータの利用技術のカリキュラムを実施するにあたり、情報処理技術の活用による個別指導の可能性を明らかにすることとした。

本研究では、利用技術の一つである Web の情報活用技術[*1]を取り上げた学生実験カリキュラムを対象に、多数の学生を同時に演習指導するにあたり、IT を活用することで個別指導の充実を図る方策を探ることを目的とする。

本稿では実験のカリキュラム、実験システム、実験支援システムを通して得られた学生の学生実験に対する取り組み姿勢、行動特性の結果を示し、IT を利用したシステムの可能性について考察する。最後に、個別指導に求められる実験システムと運用方法について提言する。

2　学生実験のカリキュラム

2.1　学生実験の目的と内容

膨大な文書を分析して内容を瞬時に把握したい、という産業界を代表する業界のニーズに応える技術、テキストマイニングの基礎手法を学び、その分析結果の記述や視覚化表現をレポートで行わせることでレポート作成を習得することを目的としている。

発展途上の技術であるテキストマイニングに問題意識を持って取り組ませることで、情報活用の可能性を体験させる。加えて、分析作業は個人の価値観に基づく作業が重要となり、学生実験の評価は「メンタリティ」「独創性」「個性」を重んじることが鍵となる。そのためには個別指導の充実が必要となり、学生個人を知り、個性を磨かせる教育が必要となる。

学生実験の具体的な内容は、

1. 茶筌[*2]とエクセルを用いたテキストマイニング
2. テキストマイニング専用ツールである「Knowledgeocean[*3]（以下 KO）」を用いたテキストマイニング

の２つの方法で、テキストマイニングの基礎技術を習得するカリキュラムになっている。

学生実験は５人一組で１班とし、次の５つの点を学生に留意させて実験を進めた。

1. 分析作業を進める前に、「どのような目的、狙いで分析するのか」を明確にし、各班で共有すること、課題を明確にすること、テーマによっては仮説を立て、自らの視点を持ち進めること
2. 目的と課題を共有後、分析手順を作成する。大雑把に１回分析し、その結果を以って班で議論し、分析方法・手順に新たな問題がないかを整理し、分析手順を見直すこと。
3. 分析にあたっては、同一データに対して、出来る限り多面的な分析を試み、複数の結果を比較検討すること。
4. 複数人で分析し、その結果を班全員で議論して結論を出すこと、結論をまとめるには、分

[*1] Google などの検索の他、情報の収集分析、テキストマイニング etc

[*2] ©奈良先端科学技術大学院大学情報科学研究科自然言語処理学講座

[*3] ©NTTDATA Knowledge

析レベルの高さ、斬新さがあること。

5. 分析過程で得られる成果物（表、グラフ、分析結果、議論内容など）は、分析結果を検証する際に必要であるのでデータとして保存しておくこと。

2.2 実施スケジュール

学生実験はコンピュータサイエンス学部２年次の学生を対象に４月２０日から７月１３日の毎週火曜日３時限から５時限（＝４．５時間)、全１２回で行なわれた。６回をエクセル、茶筌を用いたテキストマイニング、２回をプレゼンテーション資料作成と発表、４回を KO を用いたテキストマイニングに充てた。

2.3 運用

学生実験を履修した学生１５９名に対し、講師２名、ＴＡ (Teaching Assistant) ８名、担当教員３名、アシスタント１名で指導にあたった。

出席確認を行い、講師による実験概要説明後、実験を開始する。実験操作の支援を運営側で行いつつ、ＴＡは前週の実験レポートの形式評価を行い、要件が満たされていない場合は学生に再提出を指示する。

実験に必要なデータは学生実験支援 Web サイト予め用意し、学生自身にダウンロードさせた。

2.4 レポート

エクセルにデータを保存しつつ学生実験を進めさせ、レポートは各個人で作成させ、毎週エクセルファイル[*4]で提出させた。

提出は次週の実験開始時刻までに FTP により行なうこととした。

[*4] プレゼンテーションのパワーポイントファイルという例外あり

3 学生実験支援システム

3.1 システムの目的

本実験では学生実験システムを IT 化することで、個別指導の充実を行うことが目的である。そのため、コンピュータに任せて省略出来ることは任せ、人手を省略した。その上で、学生の実験に関わる全ての行動を把握・収集・分析し、個別指導の充実を行い、能力向上を目的とする。図１に能力向上プロセス示す。

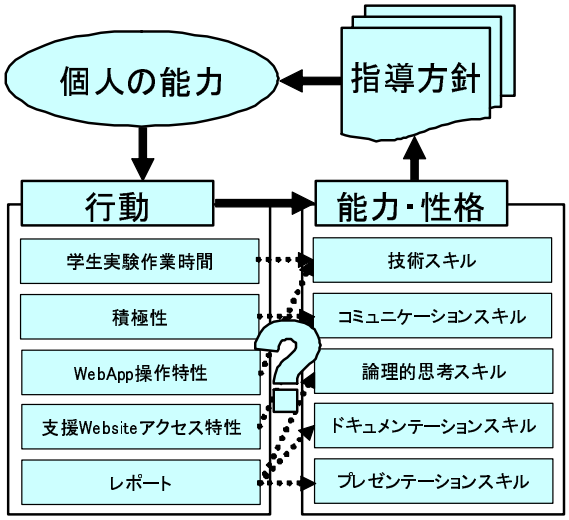

図１　能力向上プロセス

3.2 システム構成

学生個人個人が実験にて行った行動を把握するデータを収集するために、次の５つの機能を学生実験支援システムに実装した。システム構成を図２に示す。

1. レポート管理システム
2. 学生実験支援 Web サイト
3. テキストマイニングアプリケーション
4. Log 収集・分析システム
5. 世代バックアップ

WebServer 上でレポート管理やログの分析を行い、学生実験時間中はアシスタントのノートパソコ

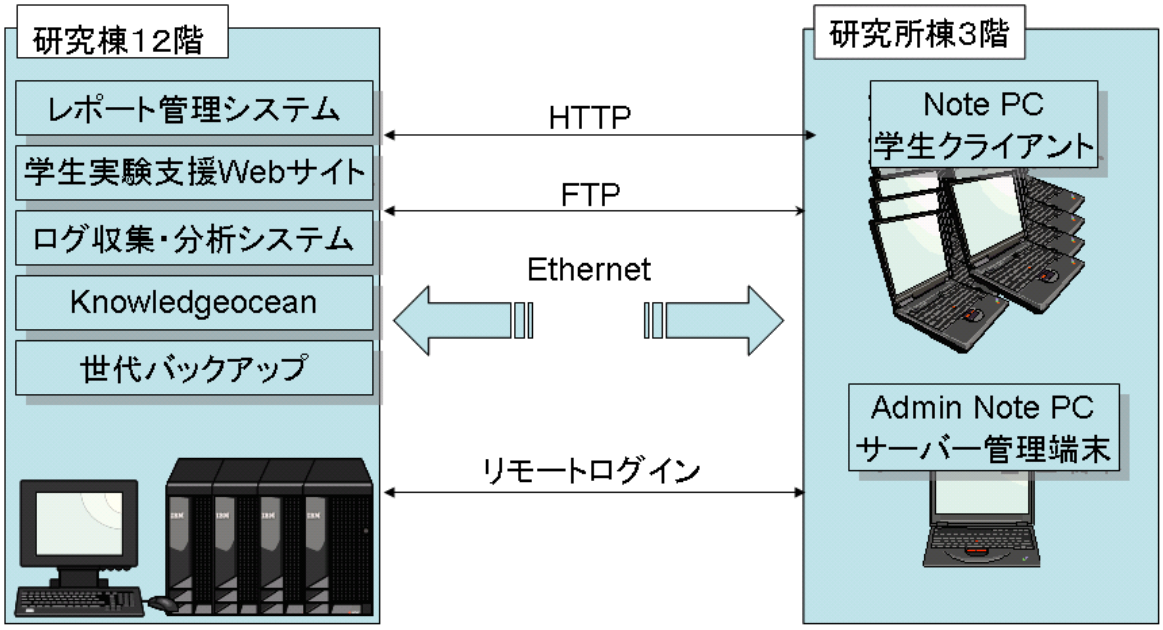

図２　システム構成

ンからリモートアクセスし、学生のアクセス状況やサーバー負荷をモニタした。

3.2.1　レポート管理システム

電子化し、レポート提出時間や場所の制約を外すことで、実験レポート作成に関する個人のワークスタイルを把握できるようにした。そのため、レポートを受け付ける FTP サーバーは学生実験時間を除き２４時間稼動させた。

学生実験時間中に FTP サーバを停止させる理由は、学生には実験に専念してもらうことと、レポート提出期限の遵守を促すためである。

提出日、遅れ提出日、再レポート提出日といった情報をＣＳＶ形式で蓄積することにより、学生個人のレポート提出状況やその傾向分析を即座に可能とした。

ＴＡや実験講師は必要に応じ何時でも何処からでも Web アクセスでレポートを閲覧できるので、学生の個別指導が容易に行なえる。

3.2.2　学生実験支援 Web サイト

情報周知やデータのダウンロード、講義資料の公開や情報交換のための掲示板を設置した支援 Web サイトを構築した。学生が実験作業を残し、帰宅してしまっていた状態で補足事項が発生した場合などには学生実験時間中に支援 Web サイトを更新し、情報の周知を徹底した。ネットワークの利用制限から当該 Web サイトへのアクセスは学内に制限した。

3.2.3　テキストマイニングアプリケーション

学生はサーバに実装された KO をアクセスし、テキスト分析を行う。KO は学生実験時間外も学内のどこからでも利用可能とした。

KO の主要機能は次の通りである。

1. 主要語抽出
   単語を抽出する
2. 概念抽出
   ”何をどうした”といった概念を抽出する
3. 属性概念
   CSV 形式のテキストデータを用いることに

より、コンセプトに対して何に属するコンセプトなのか分類出来る

4. 共起分析
   抽出したコンセプトの組を分析する
5. クラスタリング
   コンセプトを元に、類似文書をグループ分けする。いくつのグループに分けるかは１～１００で設定可能。
6. 辞書編集
   同義語・削除語・主要語などを登録することで、所望の単語を主要語抽出あるいは概念抽出できるようにする
7. テキストインポート
   分析対象となるテキストデータ（.txt と.csv のみ）をサーバーにアップロードする

3.2.4 Log 収集・分析システム

レポート提出の FTP ログやテキストマイニングツールのアクセスログ、学生実験支援 Web サイトのアクセスログを収集し、分析を行うことで、学生全体のマクロ傾向と学生個人個人の行動情報を抽出する。

具体的には、時間別、学生別、アクセス種別別にアクセス回数をカウントし、分析を行う。

3.2.5 世代バックアップ

毎週全データのバックアップを行い、システム障害に備えると同時に、学生によって上書きされていた情報の閲覧を可能とした。特に、学生実験では再提出のレポートを FTP で上書きすることで提出させていたため、修正前のデータは最新の FTP ディレクトリには残っていない。レポート提出期限直後にバックアップと取っているため、過去のバックアップのファイル存在を見ることで提出状況を把握可能である。

4 分析結果

IT を利用することで学生個人の個性把握の可能性があるのか、分析により探った。

4.1 学生実験支援 Web サイトへのアクセス状況

学生の情報収集を行なう姿勢を分析した。

学生実験支援 Web サイトでは、HTTP サーバーのアクセスログを解析することにより、学生のサイトアクセス状況を分析出来る。日別アクセス数の推移を図 3 に示す。

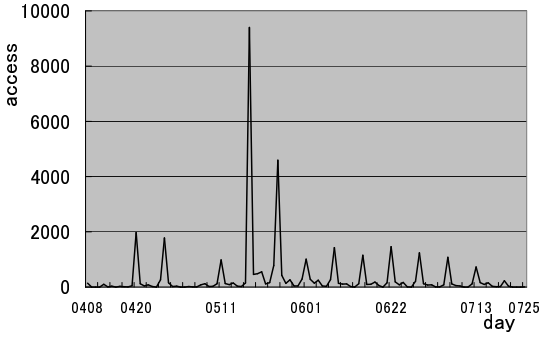

図 3　サイトアクセス状況

４月８日から７月２５日までの総アクセス数は 36,858 回であった。学生一人当たり平均 231 回アクセスしたことになる。また、実験当日のアクセスが 26,833 回で、72.8% 実験当日アクセスしている。学生実験日[*5]の平均アクセス数は 1,283 回である。

グラフから、実験当日のアクセス数が大幅に多いことは当然の結果として読み取れる。しかし、実験当日以外に平均 103 回のアクセスがあり、月曜日の平均アクセス数は 194 回、金曜日の平均アクセスが 158 回と増加傾向にあることがグラフより判明した。

これらの結果より、学生は情報が掲載されているならばある程度のアクセスは行うが、実験当日や休日前後に確認する傾向にあると考えられる。

[*5] 突出した２回のアクセスは除く

また、土日のアクセスが 0 ではないことがデータより判明した。本学では 2 年次より VPN が利用可能となり、アカウントを持つ学生は学外からでも学内に VPN を活用し、学習していることが分かった。

4.2　再レポート率の推移

IT 化実験カリキュラムに対する学生の習熟度向上を、レポート再提出率の推移により分析した。レポート再提出の判断は、学生実験のレポート作成が初めてということから、指定された形式を満たしていることを基準とし、一つでもミスがあった場合は再提出を指示した。図 4 にレポート再提出率を示す。

第４週はレポートの様式を変更したため、極端にレポート再提出率が高いが、第１週から第２週よりも、第４週から第５週にかけてレポート再提出率が低下していることから、習熟度向上によって新たな様式に対応する能力が上昇していると言える。

最終実験のレポートは全員受理としたため、参考データにはならない。

4.3　レポート提出特性

4.3.1　曜日別分析結果

図 5 に曜日別レポート提出数を示す。

レポートの期限日である火曜日の提出が 78.9% を占め、火曜日のみに提出している学生は 20.1% 存在した。土日のレポートを提出している学生もあり、次に挙げる時間帯別分析の結果、深夜や明け方の時間帯の提出が大半を占め、日中の提出は少ないことが判明した。

これにより、火曜日のみ・中間・土日経験ありの大きな３つの特性を抽出した。

4.3.2　火曜日の分析結果

図 6 に学生実験日である火曜日の１時間毎のレポート提出数を示す。

グラフからは９時台の方が提出数が、学生実験レポート提出期限である１３時台よりも多いことが読み取れるが、データを 10 分単位で集計を行った場合、期限時刻である 13:20 分の直前の提出数が、９時台の 10 分単位の平均提出数である 120 を大幅に上回る 315 であることが分析出来た。

曜日別分析と火曜日の時間帯別分析の結果より、

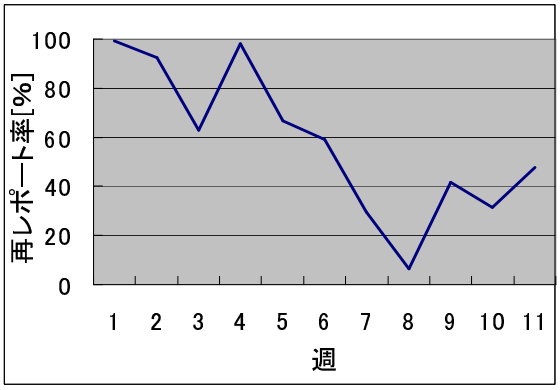

図 4　再レポート率

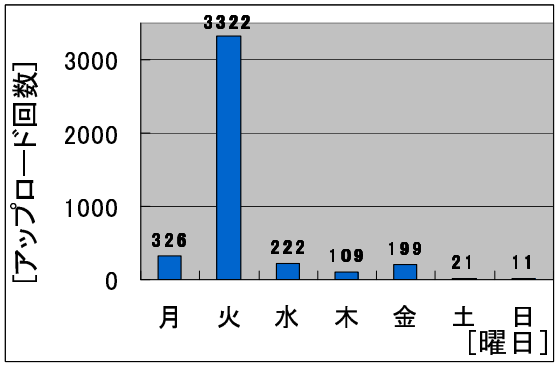

図 5　曜日別レポート提出数

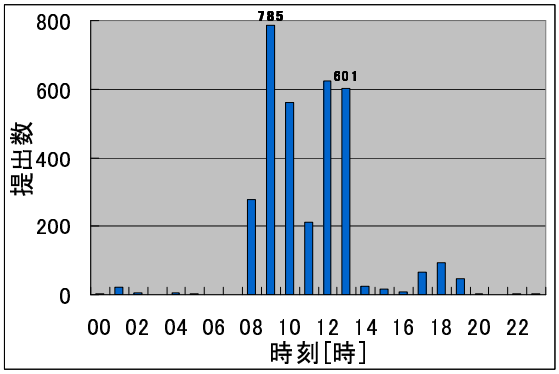

図 6　時間帯別レポート提出数（１時間）

学生はレポート期限日に提出する傾向がかなり高く、その内訳は大学到着時と期限時刻直前の２つに分けることが出来ると考えられる。その結果、レポート提出特性から５つのクラスタを作成可能であると分析した。

4.4　各種特性と成績の関係

4.4.1　KO アクセス回数

図 7 に KO アクセス回数分布を示す。

KO のアクセス回数は 12.5% が学生実験時間外のアクセスとなっている。学生別にアクセス回数を見ると、76.7% が時間外になんらかのアクセスを行っていた。

KO アクセス回数と成績との関係は、相関係数が 0.18 となったため、相関があるとは言えない。これは、分析作業は回数をこなせばいいというものではなく、要領の良い学生は少ないアクセスで良い成績を取っているからである、と思われる。

4.4.2　KO 辞書の書き換え回数

辞書の書き換え回数は 11.5% が学生実験時間外のアクセスとなっている。学生別にアクセス回数を見ると、43.4% が時間外に辞書更新を行っていた。

図 8 に KO 辞書の書き換え回数と KO アクセス相関を示す。

相関係数は 0.53 となり、ある程度高い相関を得ている。また、辞書更新回数は多く KO アクセス回数は少ない、辞書更新回数は少なく KO アクセス回数は多いという逆の特徴を持つ学生が存在していることが興味深い。

図 9 に KO 辞書書き換え回数と成績の相関を示す。

相関係数は 0.23 で、相関は高くはないがあると言える。これは KO アクセスによる分析とは異なり、日本語解析処理の特性上、分析対象に合わせた辞書の編集が良い分析には必要であることから、KO アクセスより相関が現れたと思われる。

4.4.3　レポート提出特性

土日に提出を行った学生の平均評価は 86 点と平均の 82 点より高く、土日の提出と成績評価に相関が見られた。

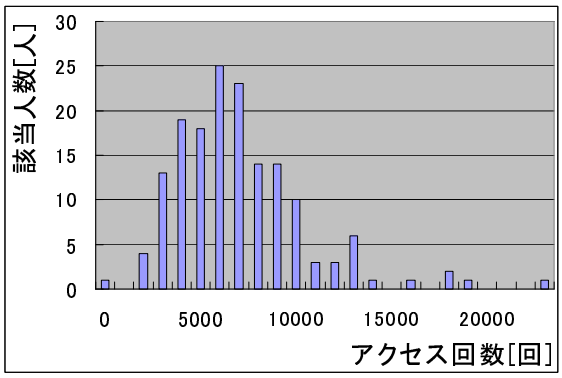

図 7　KO アクセス回数分布

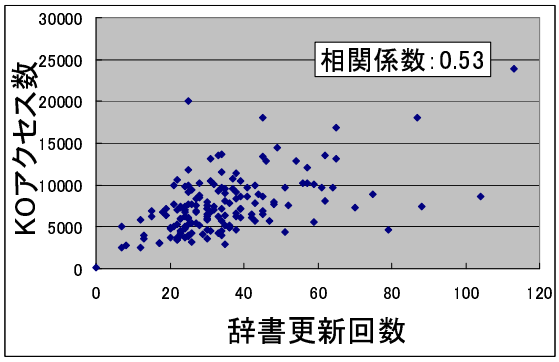

図 8　KO の辞書書き換え回数とアクセス回数相関

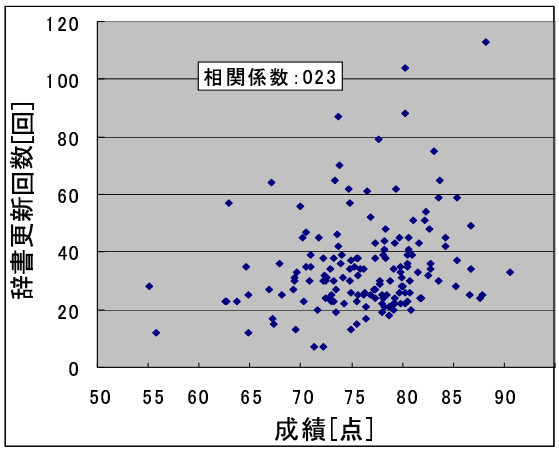

図 9　KO 辞書書き換え回数と成績の相関

5　IT 化の効果（考察）

5.1　メリット・デメリット

個人の行動を把握するために IT 化を行ったが、副産物として次のようなメリットを得ることが出来た。

- ２万枚強のＡ４用紙の節約
- レポートの整理不要
- 提出行為において、時間、場所の制約がない

逆に、紙提出と比較した場合のデメリットを挙げる。

- 機能が充足していないことによる送達確認ミス
- Face To Face の機会減少
- 印刷を意識した組版学習環境の消滅

しかし、これらのデメリットは、機能の充実、実験中の積極的なアプローチ、少々矛盾するが数回紙提出を体験させることでカバーした。

5.2　新たに分かった可能性

情報提供・収集の場から制約をなくせばなくすほど、学生個人個人の特性が出て来ることは、土日のアクセスや、学生実験当日以外に分布した学生実験支援 Web サイトへのアクセスから分かった。これらは、学生実験時間内と時間外、という２つの大きな分類によって分析した結果である。

今後は、学生実験時間中の行動を分析し、更なる個性を抽出することが可能であると考えられる。

6　今後の課題

6.1　学生実験支援 Web サイトの学生認証

学生個人の行動を把握出来るようにし、学生の個性に合わせた学生実験取り組みのコラムやレポート提出状況などの個人情報を提供する場合にはプライバシを配慮する必要があるので、学生実験支援 Web サイトに学生の個人認証機能を備える必要がある。

6.2　形式チェックのオートマチック化

実験講師が、レポートの形式チェックに稼動を使うことは、効率は悪いと言わざるを得ない。形式チェックは IT システムで対応し、実験講師は学生に FaceToFace で接し、内容に関わる指導により多くの時間を使えるようにすることが望ましい。

6.3　能力向上の評価

学生は学年ごとの特徴があり、一度構築した方法論を適用しても同様の効果を得ることは期待できない。毎年、実験カリキュラムの進行途中で、出来る限り迅速にかつ的確に学生の個性や能力を把握し、学生の個性に沿った個別指導を行い、習得効果を測定できることが必要である。

6.4　グループ編成

学生の個性・能力が把握出来たならば、個別指導だけではなく、学生同士の良い相互影響が働くように、個性の面から見たグループ編成を行う必要があると考える。