

教育用デジタルコンテンツの学習ログの分析

廣川佐千男[†] 殷 成久^{††} 島田 敬士^{††} 大久保文哉^{††} 緒方 広明^{††}

[†]九州大学情報基盤研究開発センター

^{††}九州大学基幹教育院

E-mail: [†]hirokawa@cc.kyushu-u.ac.jp

あらまし 九州大学では平成 25 年 4 月から、学生全員の PC 必携化を行っている。これにより e-Learning を含む教育の情報化が進み、教育に関する様々なデータが蓄積されている。本稿では、教育用デジタルコンテンツ配信プラットフォーム BookLooper^(注1) を利用した半期の講義について、コンテンツ配信・閲覧の利用ログに対してテキストマイニングを適用することで、学習項目や学生や利用ソフトウェアごとの学習活動の特徴と違いを抽出した。

キーワード 教育用デジタルコンテンツ, e-Learning, 学習履歴, テキスト・マイニング, ラーニング・アナリティクス

Analyzing Learning Behavior of Educational Digital Contents

Sachio HIROKAWA[†], Chengjiu YIN^{††}, Atsushi SHIMADA^{††}, Fumiya OKUBO^{††}, and Hiroaki OGATA^{††}

[†] Research Institute for Information Technology, Kyushu University

^{††} Faculty of Arts and Science, Kyushu University

E-mail: [†]hirokawa@cc.kyushu-u.ac.jp

Abstract

Key words digital contents, e-learning, learning log, text mining, learning analytics

1. はじめに

近年、Blackboard、Virtual-U、WebCT、TopClass など、Web を利用する多くの教育システムが開発されてきた。これらのシステムでは、履歴データが残るので学生の学習活動を分析することができる。ログデータには、学生の行動レコードとして、読んだページ、書いた記録、小テスト、演習やレポート課題だけでなく、他の学生との交流活動 [10] も残すことができる。学生のログを分析することで、個々の学生の学習活動の特色を把握し、教え方の改善も期待できる。

エデュケーションル・データ・マイニングとよばれる分野で活発な研究が始まっている。データ・マイニングの手法を、教育に関連するデータに適用することで教育改善を目指す研究である。[9] では、対象データに含まれる潜在的なパターンを自動的に抽出している。ログデータを分析することで、学生の特性による教育方法の効果の違いや、学生が個々の項目をどのように学習しているか、あるいは、学生がスキップしているページや項目はどこか、などを分析することができる [11]。

文部科学省は、小学校、中学、高校に利用する教科書を 2020

年までにすべて電子教科書に移行する方針を示している^(注2)。著者らが所属する九州大学では、その先行的取組として、2014 年 4 月から教育用デジタルコンテンツ システム BookLooper を利用している。これまで教科書としては本を利用していたが、本学の学生は BookLooper で講義を受講している。講義の前に質問をまとめたり、授業中に重要項目をマークしたり、メモをとったりできる。講義の後の復習にも利用できる。これらの学生の学習行動はすべて記録され、膨大な量が蓄積される。本論文では、我々が構築している学習履歴分析システムを紹介し、学習活動の分析事例を紹介する。

2. 関連研究

エデュケーションル・データ・マイニングとよばれる分野が始まっている。教育に関するデータを分析する多くの研究が始まっている。Kay ら [12] は、TRAC システムを利用して、学生のチームワーク活動における協力関係を分析している。Pechenizki ら [13] はどのような教材が学生に向いているかという推定の自動化を試みている。Blagojevic と Micic [14] は、学生の行動を予測し学習活動状況を自動的にレポートするシス

(注2): <http://www.mext.go.jp/>

テムを提案している。

本研究の分析対象とする学習ログは、BookLooperに残されたものである。BookLooperは教材提供と講義の実施に利用できるだけでなく、学生はいつでもどこでも利用して学習することができる。教員は、自身が担当しているクラスの学習活動全体を把握するだけでなく、個々の学習者についての分析も可能となる。

3. 教育用デジタルコンテンツ利用学習ログ

3.1 ログデータ

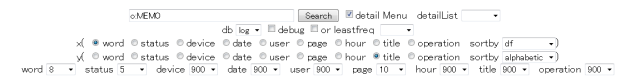
本稿では、大学1年生61名を対象として2014年4月から7月に開講された科目「情報科学」についての学習履歴を分析した。講義資料のスライドはBookLooperの資料として閲覧できる。内容としては、A-01～A-11、B-01～B-09、C-01～C-05の25項目から成り立っている。各項目はさらに数十ページのスライドで構成されている。なお、分析対象とするログは5月から7月の期間の履歴であり、その期間に行われた講義項目はBとCである。ただし、講義終盤の7月には、それまでの講義内容を振り返り重要語や疑問点をまとめる課題が出題されたので、4月、5月の講義内容であるA項目についてもアクセスされている。BookLooperでは資料は、本棚、本(項目)、ページの3段階に配置され、利用者は順次見たり、戻ったり、必要に応じてブックマークやマーカーで記録したり、メモの挿入など31種類の操作を行うことができる。本稿では、一件のログに含まれる、時刻、利用形態、利用者番号、項目、操作、ページを分析対象とした。利用形態にはAIR、FLS、AN、IPの4種類ありそれぞれPC用BookLooperアプリケーション、ブラウザ組み込みflash、Android、iPad/iPhoneアプリケーションを表している。今回分析対象としたログでは殆どがAIRとFLSだった。例えば図1は5月23日のログの一部である。

```
2014/05/23 13:06:17,AIR,user003,B-04,003,NEXT,29,2,,,
2014/05/23 13:06:17,AIR,user006,B-04,002,PREV,13,2,,,
2014/05/23 13:06:17,FLS,user004,B-04,014,ZOOM,1,2,,,
2014/05/23 13:06:17,FLS,user005,B-04,003,NEXT,5,2,,,
2014/05/23 13:06:17,AIR,user002,B-04,014,ZOOM,1,2,,,
2014/05/23 13:06:17,FLS,user001,B-04,003,NEXT,29,2,,,
2014/05/23 13:06:17,FLS,user004,B-04,013,ZOOM,1,2,,,
2014/05/23 13:06:17,FLS,user004,B-04,013,ZOOM,1,2,,,
2014/05/23 13:06:17,AIR,user007,B-04,003,NEXT,29,2,,,
```

図1 ログデータ

4. クロス表システム

5月から7月までのログ101,086件を対象として2種類の検索エンジンを作った。いずれも前節で述べた操作ログを検索対象の文書とした。各操作について、ユーザーID、利用ソフト、項目、ページ、操作はそれぞれを表す記号の前に識別のための一文字と区切記号としてコロン(:)を付加した単語を検索インデックスとして登録した。コメントは形態素解析し、名詞、動詞、形容詞をインデックスとした。



	わかる (55/2817)	学ぶ (30/30)	復習 (23/24)	しる (2/2)	れる (21/21)	知る (11/11)	ある (15/15)	分かる (13/13)
A-01 (55/2817)	5	5		2	6	11		
A-06-07-08 (38/3122)	2	1	6	2		1	2	6
A-08 (1/538)								
A-09 (21/2402)	7	5	4	3			4	3
A-10 (12/1983)	5				5	5	5	
A-11 (14/1916)	1	1		2	1			1
B-01 (24/9186)	4	1		4	3			
B-04 (17/9221)	5	1	2	2	1			1
B-05 (18/9185)	5	3			1			
B-06 (33/10617)	14		7		2			1
B-09 (9/9907)	5	1						

図2 クロス表検索エンジン

	わかる (55/2817)	学ぶ (30/30)	復習 (23/24)	しる (2/2)	れる (21/21)	知る (11/11)	ある (15/15)	分かる (13/13)
A-01 (55/2817)	5	5		2	6	11		
A-06-07-08 (38/3122)	2	1	6	2		1	2	6
A-08 (1/538)								
A-09 (21/2402)	7	5	4	3			4	3
A-10 (12/1983)	5				5	5	5	
A-11 (14/1916)	1	1		2	1			1
B-01 (24/9186)	4	1		4	3			
B-04 (17/9221)	5	1	2	2	1			1
B-05 (18/9185)	5	3			1			
B-06 (33/10617)	14		7		2			1
B-09 (9/9907)	5	1						

図3 項目と日付によるクロス表

図2はその一つであるクロス表検索システムである。入力したキーワードに対し検索を行う。検索結果の一覧は、横軸と縦軸の二つの視点を指定することでクロス表として表示される。図2では、メモの操作を含むログを検索するために検索語としてo:MEMOを入力し、結果の338件を表示している。横軸にコメント中の単語、縦軸にA-01からC-04までの項目を指定してクロス表として表示している。ただし、横軸の単語としては、検索結果のログに含まれる単語の出現頻度(df=document frequency)の上位8個に限定している。縦軸の項目は辞書式順序で上位900個と指定しているが、実際には15個しかない。また、各セルにはその列の単語とその行の項目を含むログの件数を表示している。件数が多いセルは赤色で強調表示している。

図3は横軸を日付、縦軸を項目とした結果である。日付の範囲は、ログの最初の日付5月16日から7月30日までである。5月16日にB-01が閲覧されているのが分る。それから一週間間隔で新しい項目に進んでいる。なお、この表示では最小頻度を1000にしている。これを1とすると、一週間ごとの講義の間にも他の項目がアクセスされていることが分る。C-05まで進んだ後7月18日に、A-01からC-05まで色々な項目がアクセスされている。これは、その日にそれまでの復習が指示されたことによる。

図4はログに最も多く出現した学生IDを検索条件とし、最小頻度を1とした図である。日付が進につれて項目も進んでいるのは全体の図3と同様であるが、7月18日の復習では、A-05とC-05だけを主に見ている。

5. 項目と操作の対応分析

5.1 項目ごとの操作

講義の最後の2回では、学んできた内容を各自復習する。7月18日にその指示があった。5月以前にあった項目Aについ

	0150 (01/28/77)	0152 (01/27/85)	0153 (02/13/77)	0154 (02/06/86)	0157 (01/25/81)	0160 (02/17/88)	0164 (02/13/82)	0166 (02/29/83)	0173 (02/11/83)	0175 (02/22/83)	0177 (1/7)	0178 (02/20/81)
A-A0 (01/28/77)	1								14			7
A-A0 (02/17/83)	1								25			82
A-A0 (01/28/77)	1								38			
A-A0 (01/28/83)	1											
A-A1 (0/7/84)	1											
A-E0 (01/29/80)	32	1	21									
A-E0 (01/29/81)		81	2		15							
A-E0 (01/29/85)												
A-E0 (01/29/77)												
A-E0 (01/29/87)			8	186								
A-E0 (01/29/83)						41						48
A-E0 (01/29/85)						132						34
A-E0 (01/29/78)						24	168	2				12
A-E0 (02/19/80)							138				81	
A-E0 (02/21/80)							22	22	82	104		
A-E0 (01/29/82)											12	

図 4 利用頻度上位学生の行動

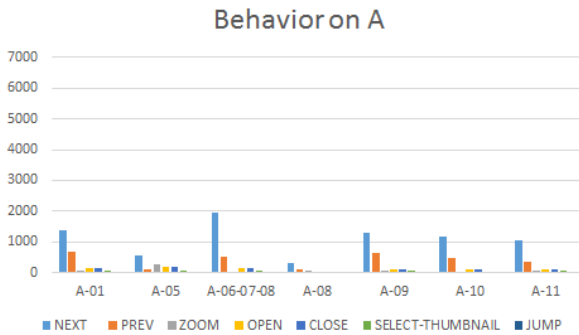


図 5 項目 A についての行動

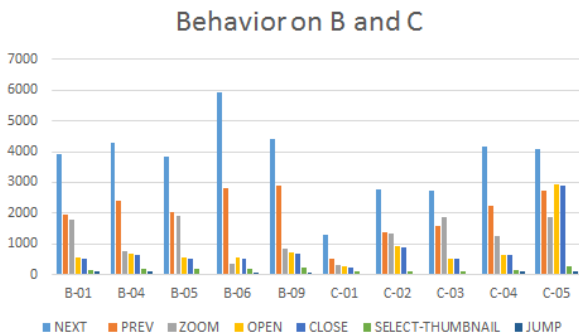


図 6 項目 B、項目 C についての行動

てもこの時点で見直すことになり、その結果、ログに残っている。図 5、図 6 は全てのログについて、項目と操作を縦軸、横軸としてクロス表を作成し、それを棒グラフとして表したものである。

項目 A は、ログ開始時期の 5 月より前に講義があったものなので、項目 B、C に比べ件数が少ない。項目 B でも項目 C でも操作として NEXT が一番多く、次に PREV が多い。ところが、項目 B では PREV が NEXT の半数以上あるのに、項目 C ではその差が小さくなっている。また、項目 C では OPEN、ZOOM、CLOSE などの操作が多くなっている。特に、最後の項目 C-05 については、C-05 の中のページを順番に見るだけでなく、JUMP が多用されていることが分る。

5.2 項目についてのコメント

7 月 18 日以降の復習の段階で、どの項目を見直し、どのようなメモを残しているか分析した。図 7、8 は各項目についてのメモ操作の頻度を棒グラフで表した。項目 A-01、A-06-07-08、C-04 についてのマーカー追加が他を抜いて多いことが分る。ただし、この分布だけでは具体的に何に着目しているかまでは分

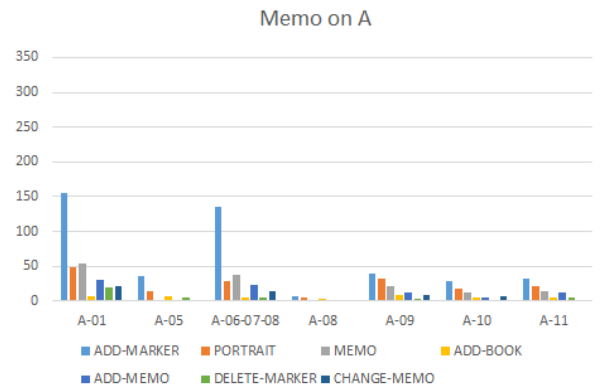


図 7 項目 A についてもメモ操作

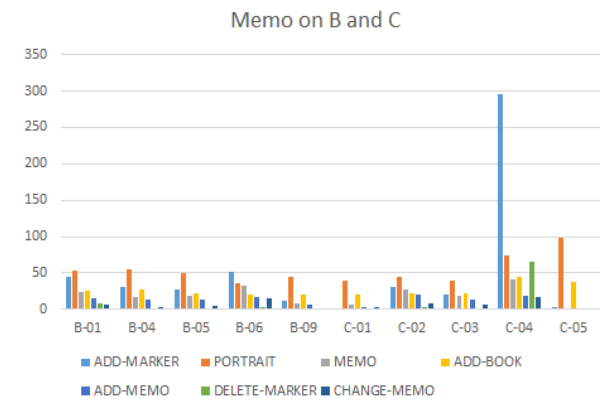


図 8 MEMO Actions on Items B ad C

らない。次の節では、内容まで解釈するため、記入されたメモに出現する単語を軸として分析する。

6. 閲覧ソフトによるコメントの違い

学生がどのようなメモを残したか、また、コメントの違いがあればその違いの要因は何かを分析するため、縦軸をコメント中の単語、横軸を色々かえて分析した。図 10 は、横軸を閲覧ソフトとしたものである。単語は頻度に降順で表示している。FLS(2 列目)と AIR(3 列目)で幾つかの単語について、出現頻度に違いがあることが分る。同じ単語について AIR 現数が全体の何割になるかという比率の降順で単語を並べたものが図 9 である。左側の単語が AIR 的な単語で、右側が FLS に特徴的な単語である。これを見ると、AIR では講義項目の個別内容の単語が多く、一方、FLS では「楽しい」「画像」「にくい」「デメリット」などの感覚的な単語が多い。

左側の単語が AIR 的な単語で、右側が FLS に特徴的な単語である。これを見ると、AIR では講義項目の個別内容の単語が多く、一方、FLS では「楽しい」「画像」「にくい」「デメリット」などの感覚的な単語が多い。

6.1 コメント中の単語のマインドマップ

どの項目のどのページについてどんなキーワードがコメント中に含まれるかを、分析する MindMap システムを [3] と同様に構築した。このシステムは、まず与えられた条件を満す文書

	vFLS (11460/49003)	vAIR (10496/59473)
わかる (39/69)	24	15
する (14/30)	7	7
思う (13/13)	11	2
れる (12/21)	3	9
いる (12/24)	7	5
方法 (11/11)	2	9
意味 (11/24)	4	7
知る (10/17)	1	9
証明 (9/9)		9
符号 (9/11)	2	7
覚える (8/10)	8	
分かる (8/13)	3	5
ある (8/15)	1	7
できる (7/9)	3	4
可能 (7/9)		7
ノイズ (7/10)	2	5
やすい (6/6)	4	2
すごい (6/6)	5	1
なる (6/8)	3	3
一意 (6/8)		6

図 9 AIR と FLS における単語頻度

群を求め、検索結果に含まれる特徴語を指定された重用度に応じて指定された個数だけ求め、単語の共起関係を MindMap として描く。ただし、全ての共起関係を表示したら複雑な図形となるので、重要度の高い単語から始め、各単語に最も類似している単語に枝を延す。こうして展開木を構成する。

本稿では、コメントに関するログを抽出するため、まず、o:MEMO というキーワードでログを検索し、得られたログに現れた単語、項目名、ページ番号、閲覧ソフトなどを特徴語として抽出した。図 11 と図 12 が求めたグラフである。赤いノードが閲覧ソフト (AIR、FLS) を表し、灰色のノードが項目、ピンクのノードがページ、色がついていないノードが単語である。

6.2 専用アプリ (AIR) 対 Web ブラウザ (FLS)

図 10 から、AIR と FLS で少数ではあるが学生のコメントが異なることが分った。具体的にどのようなコメントがあるか分析するために、単語スパニングツリーを描いた (図 12、図 11)。なお、この図では、単語の共起関連だけでなく、項目名とページも合せて表示しているの、どの項目のどのページの何につ

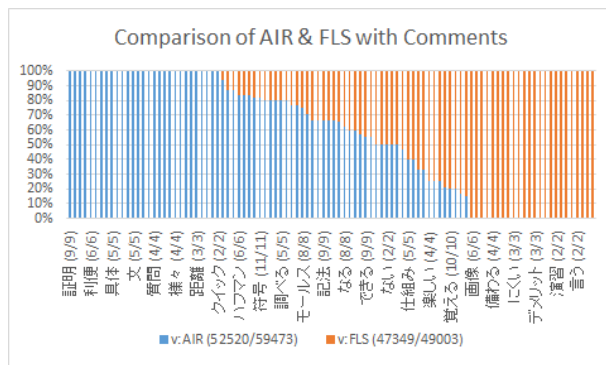


図 10 利用ソフトによるコメントの違い

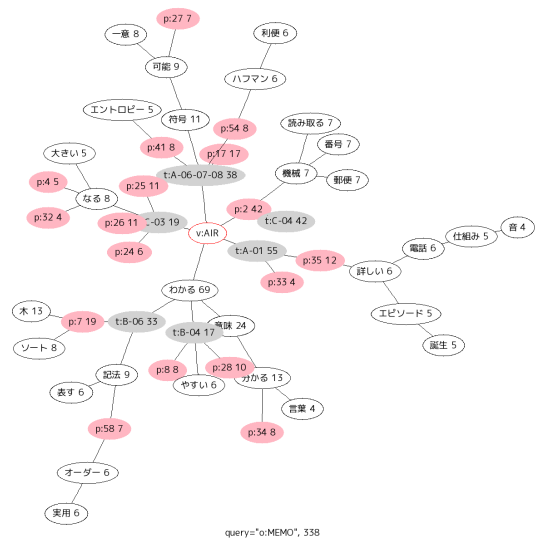


図 11 AIR 利用者のマインドマップ

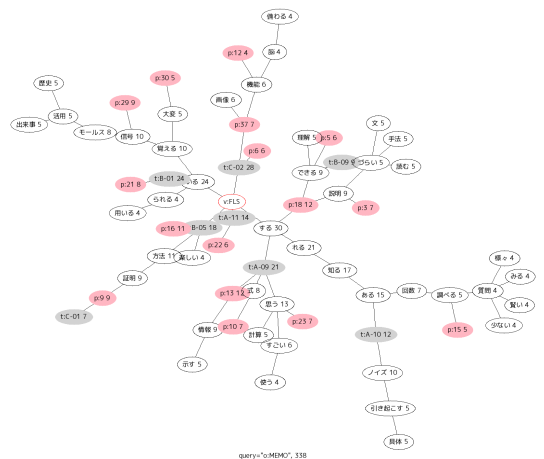


図 12 FLS 利用者のマインドマップ

いてのコメントかが分る。

専用アプリケーションで閲覧した学生についての図 11(AIR) では、各項目の内容を表す単語が多数見られる。A-06-07-08 では「ハフマン符号」、C-04 では「郵便番号読取」、A-01 では「電話の仕組み」と「誕生エピソード」、B-06 では「ソート・アルゴリズム」についてのコメントが述べられていると推測できる。図 11 で見られる主観的な表現は、B-04 についての「やすい」と A-01 の「詳しい」の二つだけである。

一方、ブラウザの flash で閲覧した学生の図 12 では、多くの主観的な単語 (形容詞) が多く見られた。A-09 では「すごい」と「思う」、B-05 では「楽しい」、B-09 では「づらい」などが特徴的である。このような形容詞は専用アプリケーションで閲覧した利用者の図 11(AIR) にはみられない。

7. まとめと今後の課題

本論文では教育用デジタルコンテンツシステム BookLooper 利用者の学習履歴の分析システムを提案した。BookLooper ではそれぞれの講義は、複数のページからなる一つの項目として構成されている。利用者はページを閲覧し、次のページや前の

ページへの移動だけでなくコメントを残すことができる。学習履歴には、項目 ID、ページ番号、操作、コメント、利用した閲覧ソフト、時刻が学生 ID とともに残される。今回の分析では学生 ID は同一利用者の識別にだけ使うため、符合化した ID を利用した。どの項目が多く見られたか、どのページを見ているときにはどのような操作が行われていたかなど、複数の観点からの分析をするため、クロス表検索エンジンを実装し分析を行った。特に、閲覧に利用するソフトにより、学生のコメントがどのように違うかを分析した。その結果、Web ブラウザ埋め込みの flash による閲覧をした利用者には、専用閲覧ソフトによる利用者には主観的なキーワードが多いことが確認できた。

短期間の小規模のデータだったが、構築したクロス表ならびにマインドマップシステムにより効率よく分析ができた。試行錯誤が必要だが、分析のための仮説設置に有効であることが確認できた。

科目による違いや、一つの科目でもページによる違い、そして、予習をしている学生とそうでない学生での行動の違いなど、まず、課題の洗い出しを行い、具体的な分析を行う予定である。

8. 謝 辞

本研究は JSPS 科研費 24500176 の助成、ならびに NICT 委託研究 178A03^(注3) による。

文 献

- [1] Ministry of Education, On Vision of Educational Computerization, http://www.mext.go.jp/b_menu/houdou/23/04/1305484.htm (accessed 2014.10.14)
- [2] Kazumasa Goda, Sachio Hirokawa, Tsunenori Mine, Automated Evaluation of Student Comments on Their Learning Behavior, Proc. ICWL2013, pp.131-140, 2013
- [3] Sachio Hirokawa, Brendan Flanagan, Takahiko Suzuki, Chengjiu Yin, Learning Winespeak from Mind Map of Wine Blogs, Proc. HIMI 2014, Part II, LNCS 8522, pp. 383-393, 2014
- [4] <http://moocs.com/>
- [5] Special Section “Grand Challenges in Science Education”, Science Vol.3, No.40, 2013
- [6] Shaymaa Sorour, Tsunenori Mine, Kazumasa Goda, Sachio Hirokawa, Comments data mining for evaluating student’s performance, Proc. AAI2014, pp.25-30, 2014
- [7] Shaymaa Sorour, Tsunenori Mine, Kazumasa Goda, Sachio Hirokawa Predicting students’ grade by measuring semantic similarity between free style comments data, Proc. ICWL2014, pp.142-151, 2014
- [8] Chengjiu Yin, Sachio Hirokawa, Jane Yin-Kim Yau, Kiyota Hashimoto, Yoshiyuki Tabata, Tetsuya Nakatoh, Research Trends with Cross Tabulation Search Engine, International Journal of Distance Education Technologies Vol. 11, No.1, pp.31-44, 2013
- [9] Klogsen, W., & Zytkow, J., Handbook of data mining and knowledge discovery. New York: Oxford University Press, 2002
- [10] Mostow, J., Some useful design tactics for mining its data, In Proceedings of the ITS2004 workshop on analyzing student-tutor interaction logs to improve educational outcomes, 2004
- [11] Romero C., Ventura S., Educational data mining: A survey from 1995 to 2005, Expert Systems with Applications, vol. 33, pp.135-146, 2007
- [12] Kay, J., Maisonneuve, N., Yacef, K., Zaiane, O, Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems (ITS 2006), pp.45-52, 2006.
- [13] Pechenizkiy, M., Calders, T., Vasilyeva, E., De Bra, P, Mining the student assessment data: Lessons drawn from a small scale case study, Educational Data Mining 2008, p.187,2008
- [14] Marija Blagojevic, Zivadin Micic, A Web-based intelligent report e-learning system using data mining techniques, Computers & Electrical Engineering. Vol. 39, No. 2, pp.465-474, 2013

(注3): <https://sites.google.com/site/9booklooper/>