Lesson Fusion System (LFS) — Applying Text Analysis to Seal the Learning Gap from One-Guideline-Multiple-Text

Hao-Min Yan[1], Ming-Hsun Yang[1,2], Shian-Hua Lin[1]*

[1] Department of Computer Science and Information Science, National Chi Nan University, Puli, Taiwan.
[2] Education Network Center of Chiayi County, Taiwan.

* Corresponding author. Tel.: +886 49 2915225; shlin@ncnu.edu.tw
Manuscript submitted November 12, 2016; accepted March 8, 2017.
doi: 10.17706/ijeeee.2017.7.1.1-12

Abstract: In Taiwan, textbooks for elementary and junior high school education are edited in accordance with the Grade 1-9 Curriculum Guidelines, then submitted to the national committee for review and approval. Various educational experts from multiple disciplines bring different perspectives into textbooks to show the advantage of content diversity based on guidelines of the curriculum. However, different versions of textbooks result in the problem of "incoherent curriculum articulation." That is the learning map of an individual student who studies textbooks from various publishers. The circumstance is worse for transfer students. Educators usually deal with this problem by focusing on the gap between different versions of textbooks and playing the tutor role in aiding transfer students. In this paper, we develop a system to automatically collect Curriculum Guidelines from three major textbook publishers (named by K, H, and N) during 2012-14 years. Using the domain "Nature and Technology" as a case study, the system extracts textbook and curriculum metadata of four grades (Grade 3-6) and merges them into the complete data set (24 textbooks: 4 grades × 2 semesters × 3 publishers). Applying text processing and mining methods, semantically same lessons from 3 publishers are firstly fused into a single class hierarchy. Then, we analyze lessons from different year-semesters to explore the life cycle of the lesson and the variation by time. Based on the result, educators can easily review the whole learning map of the Curriculum and observe the problem of incoherent curriculum articulation from the learning map as a consideration to assist students with some learning disabilities.

Key words: Text mining, curriculum articulation, catalog fusion, curriculum convergence.

1. Introduction

In Taiwan, textbooks for elementary and junior high school education are edited in accordance with the Grade 1-9 Curriculum Guidelines. Textbooks issued by different publishers will be compiled in accordance with the curriculum guidelines and submitted to the authority/agency in charge of review and approval. Regulations and requirements for reviewing textbooks are provided separately by the Ministry of Education (MOE). Various educational experts from multiple disciplines bring different perspectives into textbooks to show the advantage of content diversity based on guidelines of the curriculum. This policy is called One-Guideline-Multiple-Text (OGMT) [1]. According to local cultures and environments, schools are allowed to select adequate textbooks from all of the approved versions. However, different versions of textbooks result in the problem of "incoherent curriculum articulation." That is the learning gap occurs to an individual student who studies textbooks from various publishers during the period of Grade 1 to 9. The

situation is obvious for transfer students. Educators usually deal with this problem by focusing on the gap between different versions of textbooks and playing the tutor role in aiding transfer students. Consequently, the bottleneck is the educator workforce.

Most of the textbook authors are educational experts from various disciplines. They may have different interpretations of the curriculum guidelines so that different versions of textbooks are divergent. Therefore, learning gaps exist in the whole learning map of for individual student who studied various versions of textbooks for six or nine years. Although OGMT introduces the advantage of diversity into the learning map, the learning gap also raises from the diversity. The problem is not only troubling to transfer students but also to students who study textbooks from different publishers during two continuous semesters. Moreover, textbooks from the same publisher may have the incoherent problem due to revisions. The learning gap problem can be adequately handled by educational experts. However, they must spend time and efforts to patiently aid students.

In this paper, we develop a “Lesson Fusion System (LFS)” for comparing lessons of the “Nature and Technology” textbooks from different publishers for three years. By analyzing the metadata of each lesson, LFS automatically fuses different lesson versions that share the same or similar concepts. Finally, LFS presents the whole learning map with diversities for educators; they help students efficiently and effectively with some disabilities.

2. Related Work

Collecting textbooks from different publishers, web crawlers are developed and deployed to fetch data from sites of that publisher. The detail codes are omitted in the related work. Then, to fuse lessons of a different publisher with diverse metadata, we first refer to TW LOM as the metadata definition for lesson fusion in section 2.1. Grades 1-9 Curriculum Directory is introduced in section 2.2 to illustrated the problem derived from OGMT and how we organize lessons into the hierarchy. In section 2.3, we explore the life cycle of textbooks so that we can define the goal of lesson fusion precisely. Finally, the text analysis method referred in this paper is surveyed.

2.1. TW LOM Learning Objects

Before beginning to do the textbook lesson fusion, we must collate the curriculum-related resources to facilitate the analysis can be effective. First, we divided the curriculum-related resources into classes: Base on IEEE LOM, TW LOM completes localization needs and interoperability of learning objects. [2] TW LOM standard divide as nine categories, The granularity of learning objects is divided into Content Aggregation (CA), Sharable Content Object (SCO) and Asset. The element “life cycle” describe status and condition of SCO as a reference to discuss the life cycle of textbooks lesson. The term "Lifecycle category" describes the history and state of learning objects and the influencing factors, such as publisher, author, date of writing, etc. In this paper, to explore the lifecycle of the textbook lesson by analyzing the lifecycles and data structures of learning objects, we refer to TW LOM to collect metadata of lessons as follows.

1) Textbooks with lessons, learning contents and corresponding publishers,
2) Lessons with teaching goals, capabilities, and activities,
3) Educational materials related to each lesson, including item banks of exams.

2.2. Grades 1-9 Curriculum Directory

Grades 1-9 Curriculums are the particular framework for organizing school courses from elementary (grade 1-6) to journal high school (grade 7-9) in Taiwan. Lesson objects can be categorized from various perspectives, such as domain, grade, semester and publisher. We built the ontology of Grades 1-9 Curriculum Directory Hierarchy is shown in Fig. 1.

1) Domain Categories: Level 1 classes contain seven major domains during the elementary and junior period.
2) Grades 1-9 Categories: Level 2 classes correspond to 1-9 grades under each domain. Some domains may contain grades less than 9, for example, "Nature and Technology" consists of 3-9 grades.
3) Semester Categories: Level 3 classes are corresponding to 2 semesters each year.
4) Publisher Categories: Level 4 classes contain three publishers for each semester to denote different textbook versions.
5) Objects, Learning Objects: Finally, each leaf category contains several lessons corresponding to the version, semester, grade and domain. LFS extracts the metadata of each lesson consisting of teaching goals, activities, materials, learning objectives, exam item banks.

In this paper, we focus on "Nature and Technology" domain of elementary school. The database combines 96 themes, 834 lessons, 2,487 teaching activities, 3,708 teaching goals, 4,366 curriculum objectives and 25,397 exam items.

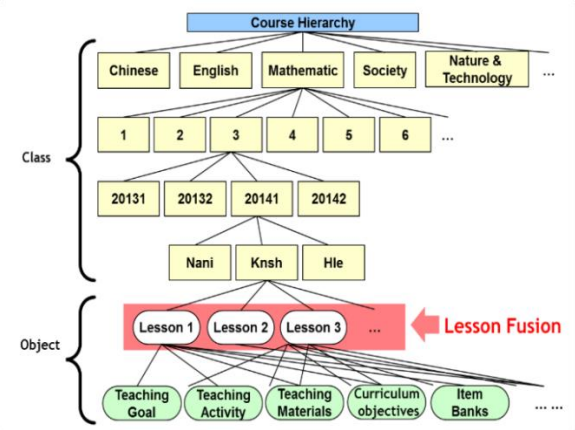

Fig. 1. Grades 1-9 curriculum directory hierarchy.

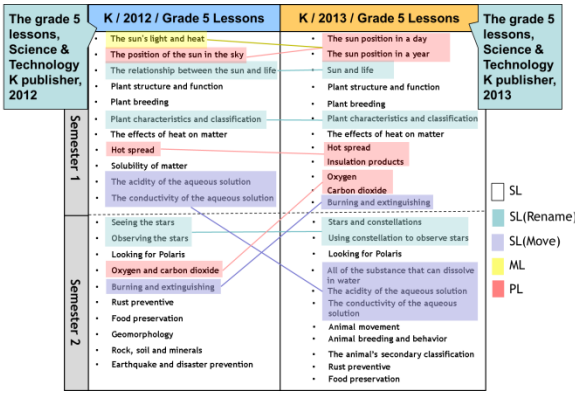

Fig. 2. The life cycle of lessons.

2.3. Life Cycle of Textbook Lessons

Publishers have to submit to be reviewed and approved by MOE for the license of publication to publish textbooks in Taiwan. The textbook license is validated for six years. During this period, the textbook content cannot be revised in the first two-year for the stabilization. At the beginning of the third year, the content can be minor revised less than half the number of pages for the trend. At the end or 6th year, the license is expired. Therefore, lessons of a textbook can be changed during its life cycle as the example shown in Fig. 2. If the license was expired between two continuous semesters, versions of the textbook are probably changes dramatically. Consequently, the status of a lesson within a textbook may be modified from the current semester/year to the next semester/year; we summarized as follows:

1) Survival Lesson (SL): A lesson continuously appears in the next semester/year, and teaching goals and activities of the lesson are not changed.
2) Merged Lesson (ML): Some lessons are merged into another lesson so that the similar teaching goals and activities are merged with few revisions in the next semester/year.
3) Partitioned Lesson (PL): A lesson is split into two or more lessons in the next semester/year. The dramatic changes of the lesson result in significant learning gaps for students.
4) Novel Lesson (NL): A lesson appears in the next semester/year, but it never appeared in previous semesters/years.
5) Removed Lesson (RL): A lesson disappears in the next semester/year. The case may result from the

revision by obeying rules defined the curriculum guidelines.

2.4. Longest Common Subsequence

The Longest Common Subsequence (LCS) [3] problem is to find the longest common subsequence between two sequences. It is often applied to applications, such as sequence alignment of molecular biology, file comparison, and screen redisplays [4]. Among the current algorithms for solving LCS problems, the systolic array algorithm [5] is most commonly used to calculate the length of the longest common subsequence for two sequences. This algorithm takes (m + n + 1) time steps to find the LCS between two sequences of length m and n, respectively.

3. Proposed Method

Lesson Fusion System (LFS) collects lesson metadata from different textbook versions in recent three years (2012-2014). Then, LFS performs text analysis process to match the same or similar lessons among textbooks from various versions and years. LFS proposes Lesson Name and Theme Fusion (LNTF) and Lesson Concept Fusion (LCF) (See Fig. 3.). The LNTF estimates pairwise similarities of name and theme between any two lessons and directly extracts similar lessons (with high scores) among different textbook versions. The LCF is then applied to extract keywords from lesson metadata and calculate conceptual similarity scores for remaining lessons.

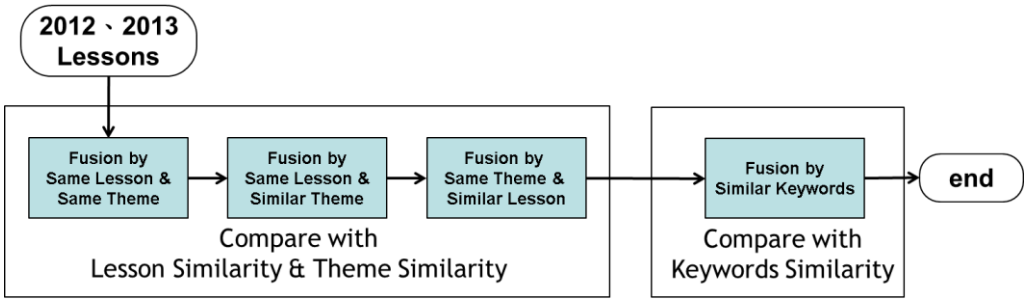

Fig. 3. The data flow of LFS.

3.1. Lesson Name and Theme Fusion

We denote lesson set of t-year as L_t, in which the set consists of lessons, i.e. $L_t = \{l_1, l_2, \dots, l_i\}$. The lesson sets of last/next year are defined as $L_{t-1} = \{l_1, l, \dots, l_j\}$ and $L_{t+1} = \{l_1, l, \dots, l_k\}$, respectively. L_t Lessons contain set elements defined as l_i and l_j within sets L_t and L_{t-1}, respectively.

LNTF calculates pairwise similarities between lessons from two set L_t and L_{t-1}. Given a lesson-pair from L_t and L_{t-1}, two metadata fields, name and theme, are individually applied to compute the LCS. Then, Jaccard Coefficient [6] are used to calculate the similar score of the lesson-pair as shown in the equation (1).

$$S_{LCS} = \frac{LCS(str_i, str_j)}{Len(str_i) + Len(str_j) - LCS(str_i, str_j)} \quad (1)$$

The str_i and str_j can be name/theme strings from lesson l_i and l_j. $LCS(str_i, str_j)$ is the estimated longest common substring of two string, and $Len(str)$ denotes the length of a string. S_{LCS} is the LCS similarity value of the lesson-pair.

LCS represents the similar part of two string from the perspective of sharing common patterns. However, the character feature of oriental languages, like Chinese, may influence the text semantics by individual characters. Therefore, matching common words between two lesson strings is also counted as the equation

(2).

$$S_{Match} = \frac{Match(str_i, str_j)}{Len(str_i) + Len(str_j) - Match(str_i, str_j)} \quad (2)$$

$Match(str_i, str_j)$ is number of words shared by two strings. S_{Match} is the word-based similarity of the lesson-pair.

Finally, two similarity measures are combined to calculate the lesson-pair similarity.

$$S_{LNTF} = \frac{S_{LCS} + S_{Match}}{2} \quad (3)$$

Lesson fusion based on LNTF must define a threshold, T_{LNTF}, to merge lessons from different lesson sets as defined in equation (4).

$$LNTF(l_i, l_j) = \begin{cases} l_i = l_j, if\ S_{LNTF} \geq T_{LNTF}\ and\ |g_i - g_j| \leq 1 \\ l_i \neq l_j, otherwise \end{cases} \quad (4)$$

The g_i and g_j represent grade of l_i and l_j. In this equation, LNTF attempts to extract the same lessons from different textbook versions, in which the delta grade value of both lessons cannot be greater than 1. In this way, LNTF is able to merge most same/similar lessons of textbooks from different publishers/semesters/years.

3.2. Lesson Concept Fusion

Comparing similarity between objects and clustering them based on keywords [7] is often proposed and often supports many applications, likes TV program recommendation [8] or studying the relationship between semantic similarity of user profile entries and the social network topology [9].

In the LCF, the system extracts a set of keywords from each lesson's metadata and then compare the same keyword numbers of two lessons. Applying Jaccard coefficient to define the similarity, the equation (5) shows the LCF similarity measure.

$$S_{LCF} = \frac{Match(K_i, K_j)}{Count(K_i) + Count(K_j) - Match(K_i, K_j)} \quad (5)$$

K_i and K_j are keyword sets extracted from lesson l_i and l_j, respectively. Similar fusion process is also applied to perform LCF by defining the threshold, T_{LCF}, as shown in equation (6). In this way, LCF is able to merge remaining lessons of textbooks from different publishers/semesters/years.

$$LCF(l_i, l_j) = \begin{cases} l_i = l_j, if\ S_{LCF} \geq T_{LCF}\ and\ |g_i - g_j| \leq 1 \\ l_i \neq l_j, otherwise \end{cases} \quad (6)$$

4. Experiments

We use the domain "Nature and Technology" as the experimental case study. Lesson counts of textbooks from publishers from 3 years are summarized in Table 1.

We invite experts (educators) of the domain to generate two answer sets: set 2012-13 and set 2013-14. By reference teaching objectives and teaching activities of the course material, experts manually perform lesson fusion and obtain results: the answer set contains 271 and 286 lessons for years 2012-13 and 2013-14, denoted as the answer sets of Y1S and Y2S, respectively.

Table 1. Lessons and Publishers of 3 Years (2012-14)

Year	K	H	N	Total
2012	91	90	92	273
2013	92	92	94	278
2014	95	94	94	283

Based on both answer sets of Y1S and Y2S, LNTF and LCF are combined to obtained two fusion set. Applying recall and precision, the F_1 measure is used to evaluate the performance.

In Lesson Name and Theme Fusion step, we have three steps to compare and fuse same/similar lessons: LF1.) Lesson fusion by same name and theme; LF2.) Lesson fusion by the same name and similar theme; LF3.) Lesson fusion by same theme and similar name.

In LF1, the system gets 223 and 217 lessons from Y1S and Y2S sets, respectively. In par with the Y1S answer set to obtain the LF1 result, the system obtains recall = 0.8229 and precision = 1.0, so that the F_1 measure = 0.9028. In the answer set Y2S, the results are recall = 0.7587, presision = 1.0 and F_1 measure = 0.8628.

Consequently, most lessons from different publisher/year can be fused by name with LF1. However, there are 50 and 55 lessons from all publishers not fused in Y1S, and there are 61 and 66 lessons in Y2S. Therefore, LF2 is then applied to fuse rest of lessons based on matching same name and similar theme. Finally, as Fig. 5, LF3 is also applied to fuse remaining lessons. Experiment results are shown in Fig. 4 and Fig. 5 present the F1-scores vary from different thresholds (0.1, 0.2, ..., 1.0). Consequently, Name and Theme Fusion achieves best F1-score with 0.94 and 0.91 for two periods. The optimal thresholds are 0.1 and 0.3, respectively.

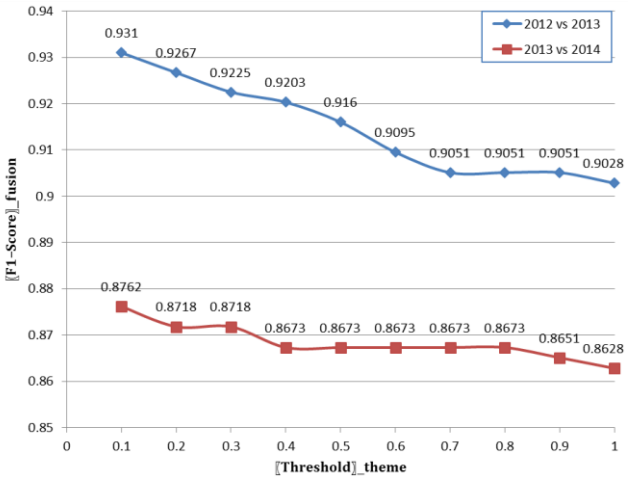

Fig. 4. Lesson fusion by same name and similar theme (LF1+LF2).

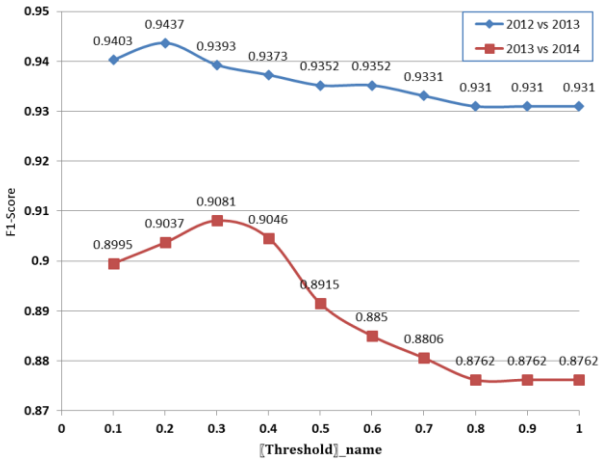

Fig. 5. Lesson fusion by same theme and similar name (LF1+LF2+LF3).

Observing the recall and precision rates shown in Fig. 4, we found that LF1+LF2 fuses 235 lessons, in par with the answer set, with lower recall = 0.8708 and perfect precision = 1 (F1 = 0.9310) in the Y1S. In the the second year set Y2S, the method fuse 223 lessons with lower recall = 0.7797 and perfect precision = 1 (F1 = 0.8762). Therefore, Name-Theme-based lesson fusion is effective but not good enough. There are remaining (37, 42) lessons in Y1S, and (55, 60) lessons in Y2S. In Fig. 5, LF1+LF2+LF3 totally fuses 240 lessons, in par with the Y1S answer set, with lower recall = 0.8856 and perfect precision = 1 (F1 = 0.9393) in Y1S. In Y2S, LF1+LF2+LF3 fuses 247 lessons with lower recall = 0.8462 and high precision = 0.9798 (F1 = 0.9081). Therefore, Name-Theme-based lesson fusion is effective but still not good enough. There are remaining (33, 38) lessons in Y1S, and (31, 36) lessons in Y2S. Rest of lessons are dramatically modified during the

lifecycle of the lesson.

The lower recall and perfect precision mean Name-Theme based fusion is efficient, but not effective, to recognize dramatically changed lessons. Finally, LCF is proposed to fuse those lessons with similar concepts. As shown in Fig. 6, the system finally obtains the fusion result with F1 scores 0.9855 and 0.9775 corresponding to Y1S and Y2S, respectively. The optimal threshold of LCF is set to 0.2.

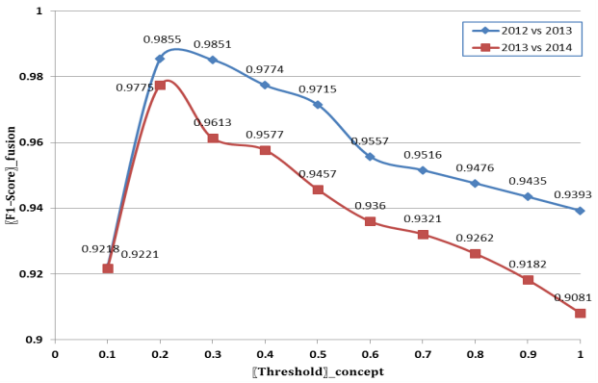

Fig. 6. Lesson fusion by similar concept (LF1+LF2+LF3+LCF).

Table 2. The Lesson Fusion Results of LFS/Experts

	SL	ML	PL	NL	RL	Orphan
K12-13	79/78	6/8	5/2	4/5	4/3	0/0
K13-14	89/81	2/4	3/1	3/3	5/3	2/3
H12-13	82/83	1/1	2/2	1/1	3/4	3/3
H13-14	86/87	0/0	3/2	3/3	7/3	2/3
N12-13	77/77	2/2	3/1	9/8	9/9	6/7
N13-14	83/74	5/7	10/4	5/7	1/8	6/6

Table 3. The Sample of Fusions

Publisher	G-S	2013			LFS Result			Remarks		Answer set by expert	
		Theme	Lesson		Theme	Lesson				Theme	Lesson
N	6-1	The weather	Water changes in morphology	6-1	The weather	Water in the atmosphere			6-1	The weather	Water in the atmosphere
			Weather and weather map			Weather and weather map					Weather and weather map
			Typhoon			Understanding typhoons					Understanding typhoons
		Oxygen and carbon dioxide	Oxygen	5-1	Air and combustion	Oxygen			5-1	Air and combustion	Oxygen
			Carbon dioxide			Carbon dioxide					Carbon dioxide
			Combustion and fire fighting			Combustion and fire fighting					Combustion and fire fighting
			Fire prevention and fire escape								
		Rusting prevention and corrosion prevention	What factors cause iron to rust	6-2	Rusting prevention and corrosion prevention	What causes iron to rust			6-2	Rusting prevention and corrosion prevention	What causes iron to rust
			A method for preventing iron from rusting			Preventing iron from rusting					Preventing iron from rusting
			Causes of food spoilage			Food spoilage and preservation					Food spoilage and preservation
			The food preservation method								
		Electromagnetic	Compass	6-1	Electricity and magnetism	Compass and geomagnetism			6-1	Electricity and magnetism	Compass and geomagnetism
			Making an electromagnet			Electromagnet					Electromagnet
						The application of electromagnets					The application of electromagnets
	6-2	Force tools	Lever	6-2	Force tools	Lever			6-2	Force tools	Lever
			The application of lever—pulley, axle			Pulley and axle					Pulley and axle
			Bicycle			Power transmission					Power transmission
				5-2	Force and movement	**Friction**		**LCF**			
		Heat in our life	The changes of heating materials	6-1	Heat in our life	The changes of heating materials			6-1	Heat in our life	The changes of heating materials
			Heat transfer			Heat transfer					Heat transfer
			Buildings in the tropics			Heat conduction					Heat conduction
						Buildings in the tropics					Buildings in the tropics
		Sustainable homeland	Development and utilization of natural resources	6-2	Sustainable homeland	~~Biology and Environment~~		**Ophan**	6-2	Sustainable homeland	Biology and environment
						Humans and environment					Humans and environment
			Green action for earth			Care for the environment					Care for the environment

Detail results shown in Table 2 explain the lesson fusion results processed by LFS and experts. In par with LFS and experts, LFS is close to the quality of experts. Partial results of fused lessons are summarized in Table 3. Theme/Lesson denoted by red color font are wrong predictions according to expert's curations. The lesson "Friction" of the second semester of grade 5 (5-2) is a wrong fusion predicted by LCF module that merges two lessons due to similar keywords from teaching objectives. In fact, both lessons are dependent that "5-1 Friction" provides concepts about friction, push, pull, ground, wheel, move, etc. for the understanding content of "6-2 Bicycle". Another wrong prediction "6-2 Biology and Environment" is corresponding to the curriculum change from 2012 to 2013, in which this lesson was added as a new lesson. Therefore, LFS predicts the object as an "Orphan" lesson due to less similar keywords.

According to the life cycle of lessons defined section 2.3, we explore the wrong results of LFS corresponding to columns 2-7. These counts of missing fusions within Table 2 are summarized in Table 4 and explained as followings.

Table 4. The Sample of Missing Fusions

Publisher	G-S	2013			LFS Result (2013-14)		Remarks		Answer set by expert (2013-14)		
		Theme	Lesson		Theme	Lesson			Theme	Lesson
K	6-2	Simple machinery	The tools that help us		6-2	Simple machinery	The tools that help us		6-2	Simple machinery	The tools that help us
							Understanding lever				Understanding lever
							The applications of lever				The applications of lever
			Gears, chains and belts				Understanding lever	PL			Power transmission
							The applications of lever				
			Fluid power				Power transmission				
N	5-2	Animal life	How do animals move?		5-2	Animal life	Animal behavior		5-2	Animal life	Animal behavior
			Foraging, avoiding the enemy and nesting								
			Survival adaptations				-	ML			Courtship and reproduction
			Courtship and reproduction				Courtship and reproduction				Animal Classification
			Animal Classification				Animal Classification				Biology and environment
	6-2	Sustainable homeland	Development and utilization of natural resources		6-2	Sustainable homeland	~~Biology and environment~~	Ophan	6-2	Sustainable homeland	Humans and environment
							Humans and environment				Care for the environment
			Green action for earth				Care for the environment				

Publisher	G-S	2012			LFS Result (2012-13)		Remarks		LFS Result (2013-14)		Remarks
H	5-2	The changes of earth's surface	Earthquake and disaster prevention		6-2	~~The changes of earth's surface~~	~~Earthquake and disaster prevention~~	RL	6-2	The changes of earth's surface	Earthquake and disaster prevention	NL
			Rocks, soil and minerals				~~Rocks, soil and minerals~~				Rocks, soil and minerals	
			The changes of landscape				~~Fluvial process~~[1,2,4]				Fluvial process	
	6-1	Animal	Animal classification		5-2	Animail world	Animal classification		5-2	Animail world	The animal's secondary classification	
			Animal movement				Animal's moves				Animal's moves	
			Animal's reproduction and behavior				Animal reproduction and behavior				Animal's reproduction and behavior	
					6-1	Animail world	The animal's secondary classification	SL			
							Animal movement				
							Animal's reproduction and behavior				

1) Survival Lessons (SL) are closed in par with LFS and experts, except for K13-14 and H13-14. More SL cases result from the dramatical changes of curriculum, in which H2012 "6-1 Theme Animal" was moved to H2013 "5-2". Therefore, extra lessons are fused into the theme.
2) Merged Lesson (ML) results contain on missing lesson in LFS "N2013-14". Three lessons are merged into one lesson "5-2 Behaviors of Animals" by experts. However, one lesson is missed by LFS due to

fewer keywords from the lesson "5-2 Survival Adaptation".

3) Partitioned Lesson (PL) presents results that some lessons are highly similar to another lesson so that they are duplicated and matched.
4) Novel Lesson (NL) results are original from the dramatical change of curriculum that "5-2 Theme Surface Change of the Earth" was moved to "6-2 Theme Surface Change of the Earth". Therefore, LFS predicts these lesson as NL in Y2S and as RL in Y1S.
5) Removed Lesson (RL) is suffered by the same reason of NL.
6) Orphan lessons mean that some lessons are not matched in the next year fusion. This case is not considered in methods above so that we define it here.

Table 5. Time Cost of Manual Curations

	Spend Time (min.)	
Publisher	**Y1S (2012-13)**	**Y2S (2013-14)**
N	34	48
K	39	42
H	41	52
Total	**114**	**142**

Table 6. Run Time of LFS on a General PC

	Spend Time (sec.)	
	Y1S (2012-13)	**Y2S (2013-14)**
LF1	0.01	0.01
LF2	3	3
LF3	7	8
LCF	65	74
LFS	**75.01**	**85.01**

Table 7. The Changes of Lesson Life Cycle

Publisher	Theme	Lesson	2012	2013	2014	Problems due to curriculum changes
H	The changes of earth's surface	Earthquake and disaster prevention	5	N/A	6	RL problem: Some removed lessons predicted by LFS are due to dramatic changes in the curriculum. For example, all lessons of the theme "The changes of earth's surface" of (Grade5, H2012) was migrated to (Grade6, H2013). Therefore, the theme of H2013 was disappeared (denoted by N/A) and started from the year 2014. That is elementary Grade5 students will learn the old curriculum of H2013 while they are grade 6. Some wrong predictions of RL result from this reason.
		Rocks, soil and minerals	5	N/A	6	
		The changes of landscape *12	5	N/A		
		Fluvial process *13,14		N/A	6	
	Animal	Animal classification *12	6	6		SL problem: Some survival lessons are predicted as correct results during the curriculum change of current/next years. However, LFS can detect the curriculum change that lessons of the theme "Animal" were exchanged between (Grade6, H2012) and (Grade5, H2014). Are correct with semantic lose. H2013 for Grade5 and Grade6 are presented as the transition period for the curriculum change. Some changes of lesson names between two-year are also detected as the * notations.
		Animal moves	6	5,6	5	
		Animal reproduction and behavior	6	5,6	5	
		The animal's secondary classification *13,14		5	5	
K	Aqueous solution	The conductivity of aqueous solution	6	5,6	5	
		The observation of dissolution *12	6	6		
		The acidity of aqueous solution	6	5,6	5	
		Dissolution *13,14		5	5	

* Names of this lesson were changed from 2012 to 2013,14 versions.

Although LFS is not better than experts, LFS is a better solution for reducing efforts of educators from ten thousands of schools in each semester. Thinking the case of "Natural and Technology" class, educators must spend the time to realize the learning gaps of transfer students. Time costs spent by experts to generate two answer set and run time of LFS are shown in Table 5 and Table 6. In Table 5, experts spent totally 114 and 142 minutes for curating two answer sets from three publishers' two-year textbooks. However, the time spends of LFS are 75.01 and 85.01 seconds without considering the time cost of collecting and

organizing. In fact, these collected data are analyzed and extracted to support our educational web services, Edu2 (Education 2.0 Platform) [10]. By reviewing the specific curriculum, we found that the curriculum was dramatically changed during 2013-2014. Therefore, experts spent more time to curate the Y2S data set and LFS also spent more time to calculate the result.

Exploring the life cycle of textbooks and problems due to curriculum changes are also an emergent topic. The result is summarized as shown in The curriculum also changes deeply influence the learning effect of students, especially for transfer students. For example, thinking transfer students using K2013 textbook version while they are grade 5, then they migrate to a new school using H2013 for grade 5 students. In Table 8, we summarize the learning gap problem due to OGMT. In the next year (2014), these students are grade 6 in the new school that follows H2014 as the textbook. Consequently, some lessons are missed while they are grade 5. Some lessons are redundant for transfer student uses, although lesson names are changes.

Therefore, LFS is able to analyze textbook data to explore these problems automatically. Finally, the system present tabular data of visual graphs for educators so that they can immediately realize the influence of curriculum changes on their students during the change of semester.

with explanations on related problems. Some error predictions of LFS are due to the curriculum change, especially for the case of lesson exchange between two grades. By considering such curriculum changes to revise those error cases, the F1-score of LFS can be improved to 0.9826 and 0.9819 corresponding to Y1S and Y2S, respectively. Consequently, LFS approximates the duration result of experts without manual efforts and can be easily deployed as web services for educators in Taiwan to aid students with learning disabilities or transfer students.

Table 8. The Curriculum Articulation Problem of the Year 2014 That Transfer Students (grade 5, K2013) Who Migrated to a School (H2013) Using H2014 Textbook Version

Grade 5		Students learn different versions of textbooks
K2013	**H2013**	K2013 vs. H2013
1. Observing the sun	1. Observing the sun	Transfer students have learned lessons A and B from K2013 but never learned C and D from H2013 while they are grade 5.
2. The plant	2. The mystery of plants	
3. Aqueous solution	6. Aqueous solution	
5. The beautiful starry sky	5. The bright starry sky	
6. Combustion and rusting	4. Air and combustion	
7. The Animal	7. The Animal	
4. Force and movement [A]	3. The effect of heat on matter[C]	
8. Sound and musical instruments [B]	8. Prevent rusting and food preservation [D]	
Grade 6		
K2014	**H2014**	**Articulation Problem due to curriculum changes**
1. The weather	1. The weather	In the new year (2014), students have to learn lessons shown in column H2014, but transfer students already learned “2. Sound and Instrument” and “5. Mechanics of daily life”. However, transfer students have learned both lessons so that educators (tutors) can teach them lessons C and D during the period of lesson 2 and 5.
2. Microorganisms and food preservation	2. Sound and Instrument	
3. The mystery of the earth	3. The changes in the earth 's surface	
4. Electromagnetic interaction	4. Electromagnetic interaction	
5. Simple machinery	5. Mechanics of daily life	
6. The effect of heat on matter	6. Simple machinery	
7. Biology and environment	7. Biology, environment and natural resources	

The curriculum also changes deeply influence the learning effect of students, especially for transfer students. For example, thinking transfer students using K2013 textbook version while they are grade 5, then they migrate to a new school using H2013 for grade 5 students. In Table 8, we summarize the learning gap problem due to OGMT. In the next year (2014), these students are grade 6 in the new school that follows H2014 as the textbook. Consequently, some lessons are missed while they are grade 5. Some lessons are redundant for transfer student uses, although lesson names are changes.

Therefore, LFS is able to analyze textbook data to explore these problems automatically. Finally, the system present tabular data of visual graphs for educators so that they can immediately realize the influence of curriculum changes on their students during the change of semester.

5. Conclusion

In this paper, LFS (Lesson Fusion System) is proposed to deal with the divergence of textbook versions due to the Grade 1-9 Curriculum Guidelines based on the goal One-Guideline-Multiple-Text (OGMT). The educational policy has the advantage of diversity, but the problem of learning gaps due to OGMT textbooks troubled students with some disabilities and their tutors. For solving the problem, metadata of lessons collected from three publishers is analyzed to present the difference between two lessons of textbooks from different publishers/semesters/years. The similarity-based method is applied to estimate the scores of both similar lessons so that similar lessons are merged, and divergent lessons are denoted in the learning map. First, Lesson Name and Theme Fusion is proposed with three modules (LF1, LF2, LF3) to fuse same/similar lessons precisely gradually. Then, Lesson Concept Fusions applied to merge remaining lessons to improve the recall rate. Consequently, LFS can effectively and efficiently fuse lessons in par with the result of manual fusion by the manpower of educators. Experiment results show that F1 scores are 0.9826 and 0.9819 in two data set.

In the future, the life cycle of lessons appearing in textbook versions is worth to analyze and to explore the educational trends. Based on the learning map with gaps from OGMT, an adaptive quiz/exam system can be effectively developed to support educators testing the learning gaps of transfer or weak-progress students. Analyzing the test result and the learning map, the system can efficiently figure out the bottleneck of learning for the student. Consequently, a web-based educational platform can be developed based on the result.

References

[1] *Ministry of Education, Publication. Grade 1-9 Curriculum Guidelines.* (2015). from the website: http://english.moe.gov.tw/public/Attachment/66618445071.doc

[2] Chen, Y. N., Chen, S. J., Cheng, C. J., Chung, F. C., & Chen, H. P. (2006). TWLOM: An application profile approach. *ICDAT 2006*.

[3] Maier, D. (1978). The complexity of some problems on subsequences and supersequences. *Journal of the ACM (JACM), 25(2), 322-336*.

[4] Gosling, J. (1981). A redisplay algorithm. *ACM SIGPLAN Notices, 16(6), 123-129*.

[5] Robert, Y., & Tchuente, M. (1985). A systolic array for the longest common subsequence problem. *Information Processing Letters, 21(4), 191-198*.

[6] Suphakit, N., Jatsada, S., Ekkachai, N., & Supachanun, W. (2013). Using of Jaccard coefficient for keywords similarity. *Lecture Notes in Engineering and Computer Science*, *2202(1)*, 380-384.

[7] Rezaei, M., & Fränti, P. (2014). Matching similarity for keyword-based clustering. *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)* (pp. 193-202). Springer Berlin Heidelberg.

[8] Wartena, C., *et al.* (2011). Keyword-based TV program recommendation. *The 9th Workshop on Intelligent Techniques for Web Personalization* (pp. 15-21).

[9] Bhattacharyya, P., Garg, A., & Wu, S. F. (2011). Analysis of user keyword similarity in online social networks. *Social Network Analysis and Mining, 1(3),* 143-158.

[10] Yan, H. M., Chang, Y. L., Zeng, W. S. & Lin, S. H. (2012). Building future e-classroom with software and hardware integrations based on web 2.0 services. *Proceedings of The Second Asian Conference on Technology in the Classroom* (pp. 236-247).

Hao-Min Yan is currently studying a Ph.D. at the Department of Computer Science and Information Science, National Chi Nan University, Puli, Taiwan. His research focuses on the topic of web directory and search engines. He received his master of image processing from the Department of Computer Science at National Pingtung University, Taiwan. His future research interests are big data and machine learning.

Ming-Hsun Yang is currently studying a Ph.D. at the Department of Computer Science and Information Science, National Chi Nan University, Puli, Taiwan. His research focuses on the topic of technology in education. He received his master of education in knowledge management from National Chi Nan University, Taiwan. He is a senior teacher in Education Network Center of Chiayi County, Taiwan now. His future research interests are big data and ICT in education.

Shian-Hua Lin received his B.S., M.S. and a Ph.D. degree in engineering Science from National Cheng-Kung University, Tainan, Taiwan, in 1992, 1994, and 2000, respectively. He is major in computer science. His research interests include "web information retrieval, extraction, and classification," machine learning, "database and knowledge management system," web social network and digital learning systems. He is an associate professor of the Department of Computer Science and Information Science, National Chi Nan University now.