مجلة جامعة تشرين للدراسات والبحوث العلمية _ سلسلة العلوم الهندسية المجلد (29) العدد (4) 2007
Tishreen University Journal for Studies and Scientific Research- Engineering Sciences Series Vol. (29) No. (4) 2007

استخدام الخوارزميات في بناء وتخزين حزمة محتوى تعليمي في مستودع كائن التعلم وفق معيارSCORM

الدكتورة كندة أبو قاسم*
الدكتور بلال شيحا**
هناء سلمى***

(تاريخ الإيداع 29 / 8 / 2007. قُبِل للنشر في 2007/10/10)

□ الملخّص □

سعينا من خلال هذا البحث إلى استخدام الخوارزميات في بناء محتوى تعليمي(Content) خاضع لمعيار SCORM(Shareable Content Object Reference Model) ومن ثم تخزينه ضمن مستودع كائن التعلم((Learning Object Repository (LOR). وقد شملت هذه العملية عدة خطوات وهي تحويل حزمة المحتوى(Content Package) المشكلة وفق معيار SCORM إلى تركيب يشبه شجرة ويدعى شجرة المحتوى (Content Tree(CT)) وذلك لتمثيل كل قطعة من المادة التعليمية بالاعتماد على مفهوم نموذج فضاء الحالة(Vector Space Model)، وعلى قانون TF×IDF(Term Frequency × Inverse Document Frequency)، وعلى مفهوم خوارزميات الأشجار، وذلك لتمثيل كل عقدة محتوى بشعاع مميزة يعبر عن معلومات هذه العقدة، ومن ثم تجميع عقد المحتوى في عناقيد باعتماد قانون قياس التشابه(similarity measure) بين كل عقدتين من عقد أشجار المحتوى ووفقاً لعتبة تشابه محددة وتم تعديل دقة العناقيد من خلال قياس التشابه للعناقيد الناتجة.

الكلمات المفتاحية:

التعلم الإلكتروني ، مستودع كائن التعلم، معيار SCORM، حزمة المحتوى، الخوارزميات، قياس التشابه.

* أستاذ مساعد في قسم هندسة الحاسبات والتحكم الآلي، كلية الهندسة الميكانيكية والكهربائية، جامعة تشرين، اللاذقية، سورية.
** مدرس في قسم هندسة الحاسبات والتحكم الآلي، كلية الهندسة الميكانيكية والكهربائية، جامعة تشرين، اللاذقية، سورية.
*** طالبة ماجستير، قسم هندسة الحاسبات والتحكم الآلي، كلية الهندسة الميكانيكية والكهربائية، جامعة تشرين، اللاذقية، سورية.

مجلة جامعة تشرين للدراسات والبحوث العلمية _ سلسلة العلوم الهندسية المجلد (29) العدد (4) 2007
Tishreen University Journal for Studies and Scientific Research- Engineering Sciences Series Vol. (29) No. (4) 2007

Using Algorithms for Building and Storing Content Package in a SCORM Compliant Learning Object Repository

Dr. Kinda Abu Kassem*
Dr. Bilal Chiha**
Hanaa Salma***

(Received 29 / 8 / 2007. Accepted 10/10/2007)

□ ABSTRACT □

This research uses algorithms for building SCORM compliant learning content, storing it within a learning object repository. This process includes many steps: transforming SCORM compliant content package into a composition similar to a tree called "Content Tree" for representing each part of learning material; depending on concept of vector space model; Term Frequency-Inverse Document Frequency (TF-IDF) weighting scheme; and concept of trees algorithms. The aim is to represent each content node by a feature vector that expresses information of this node and then gathers content nodes in clusters by adopting a similarity measure between every two nodes of tree content nodes according to a specific similarity threshold. The accuracy of clusters has been adjusted through measuring the similarity of resulted clusters.

Keywords: E-Learning, Learning Object Repository (LOR), SCORM Standard, Content package, Algorithms, similarity measure.

* **Associate Professor, Department of Computer Engineering and Automatic Control, Faculty of Mechanical and Electrical Engineering, Tishreen University, Lattakia, Syria.**

** **Assistant Professor, Department of Computer Engineering and Automatic Control, Faculty of Mechanical and Electrical Engineering, Tishreen University, Lattakia, Syria.**

*** **Postgraduate Student, Department of Computer Engineering and Automatic Control, Faculty of Mechanical and Electrical Engineering, Tishreen University, Lattakia, Syria.**

المقدمـة:

تخزن عادة المواد التعليمية في نظم التعلم الإلكترونية في قاعدة بيانات تدعى مستودع كائن التعلم(Learning Object Repository (LOR). ونظراً لاعتماد معيار SCORM على نحو واسع من قبل منظمات التعلم الإلكتروني فقد تم إنشاء وتطوير مواد تعليمية خاضعة لهذا المعيار. وأدى هذا إلى نشوء عدد ضخم من المواد التعليمية الخاضعة لهذا المعيار، والذي أدى بدوره إلى مشاكل في إدارة تلك المواد. لذا، ركزت منظمة SCORM على ابتكار الطرق لحل تلك المشكلة وعلى إمكانية التعامل مع مستودعات كائنات التعلم(LORs) بحيث تلبي حاجات المستخدمين.

ومن أجل تسهيل عملية تخزين المحتوى التعليمي الخاضع لمعيار SCORM فقد كان لا بد أولاً من بناء محتوى تعليمي منسجم مع معيار SCORM، ومن ثم تخزين هذا المحتوى في مستودع كائن التعلم (LOR). وقد شملت عملية التخزين عدة خطوات وهي تحويل حزمة المحتوى المشكلة وفق معيار SCORM إلى تركيب يشبه شجرة ويدعى شجرة المحتوى (CT)، وذلك لتمثيل كل قطعة من المادة التعليمية ومن ثم إنشاء مخطط تجميع المحتوى وفق نمط المستوي (LCCG) بالاعتماد على مفهوم نموذج فضاء الحالة على قانون TF×IDF وعلى مفهوم خوارزميات الأشجار.

هدف البحث وأهميته:

يهدف هذا البحث إلى بناء محتوى تعليمي منسجم مع معيار SCORM الذي يهتم بالمحتوى التعليمي. كما يهدف إلى وضع الأسس اللازمة لبناء هذا المحتوى ضمن مستودعات كائنات التعلم بهدف التقليل قدر المستطاع من المشاكل الناجمة عن التنوع الكبير في المواد التعليمية الخاضعة لمعيار SCORM والذي ينجم عنه سوء إدارة تلك المواد.

طريقة البحث:

قمنا في هذا البحث باتباع الخطوات التالية:

1. دراسة معيار SCORM ومفهوم المحتوى التعليمي وحزمة المحتوى من وجهة نظر هذا المعيار.
2. استعراض آلية بناء محتوى تعليمي وفق معيار SCORM لدرس مختار.
3. دراسة الخوارزميات التي تسهم في تسهيل عملية بناء المحتوى التعليمي ضمن مستودعات كائنات التعلم الخاضعة لمعيار SCORM.
4. بناء شجرة محتوى تعليمي افتراضي من حزمة، محتوى، وتوحيد عمق الشجرة، وتمثيل معلومات كل عقدة بشعاع مميزة.
5. تجميع شجرات المحتوى في عناقيد اعتماداً على معدل أشعة المميزة لعقد المحتوى.
6. حساب معدل أشعة المميزة للعناقيد (الذي يمثل مركز العنقود) لتأمين دقة العناقيد.

خطوات البحث:

1. معيار SCORM

معيار SCORM هو اختصار لـ النموذج المرجعي لكائن المحتوى القابل للتشارك وقد أطلق من قبل شبكة التعلم الموزعة المتقدمة[1][2]. يصف SCORM تطوير و تحزيم و تسليم مواد التدريب و التعليم وهو يضمن بأن المناهج الخاضعة للمعيار هي "**RAID** [2][3]أي:

- **Re-usable**: معدلة ومستخدمة من قبل أدوات تطوير مختلفة.
- **Accessible**: يمكن أن تُبحث و توفر عندما تحتاج من قبل الطلاب و مطوري المحتوى التعليمي.
- **Interoperable**: تعمل عبر تشكيلة واسعة من الأجهزة و أنظمة التشغيل و متصفحات الوب.
- **Durable**: لا تتطلب تعديلات مع النسخ الجديدة من برامج النظام.

2. المحتوى التعليمي

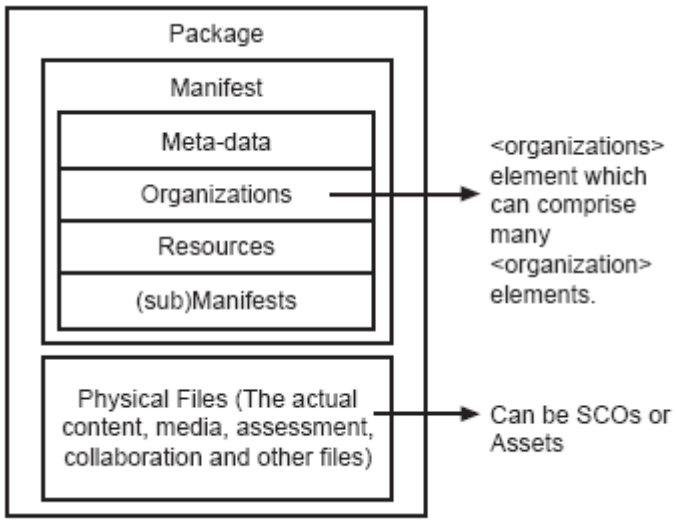
المحتوى هو نواة التعلم الإلكتروني. تعد كائنات التعلم الجزيئات الأولية لأي نظام تعلم إلكتروني، وهي عبارة عن قطع كبيرة من المعطيات المستخدمة في أنظمة التعلم إلالكتروني. ويمكن التعبير عن كائن التعلم بأنه جزء رقمي من منهج تعليمي متفاوت في الحجم والتعقيد من رسم بياني مفرد إلى المنهج الكامل.

كما أن كائنات التعلم تسير بعدّة أسماء في حقل التعلم وهي: كائنات التعليم وكائنات المعرفة وكائن التعلم القابل لإعادة الاستخدامRLO (Reusable Learning Object) وكائن المحتوى القابل للمشاركة SCO(Sharable Content Object)[2]. إن مفهوم كائن التعلم SCO الذي وضع من قبل SCORM يمثل التعبير المقبول الأكثر شيوعا فيما يتعلق بعناصر التعلم و إعادة الاستخدام. يتضمن تركيب SCO عناصر محتوى أخرى تدعى assets والتي هي تمثيلات إلكترونية من الوسائط، النص أو الصور أو الصوت أو صفحات الوب.

3. حزمة المحتوى

تحزم أغراض المحتوى في منظومة محتوى (Content Organization)، يجري توصيفها ضمن ملف خاص يدعى (Manifest)، وقد تُمثِّل حزمة المحتوى منهاجاً أو درساً أو معياراً، أو مجموعة من الأغراض المترابطة.

إن ملف Manifest عبارة عن ملف XML يدعى(imsmanifest.xml) يصف محتوى الحزمة وقد يتضمن وصفاً اختيارياً لتركيب المحتوى التعليمي (Content Structure) الشكل (1). وهو مؤلف من 4 أقسام وهي: 1) Metadata: تصف تركيب المادة التعليمية، 2) Resources: ترمز إلى المادة الفيزيائية التي تصف خاصية أو الصفة المميزة للمادة التعليمية، 3) Organizations: تحفظ الملف المرتبط بكل كائن تعلم ضمن المادة التعليمية، 4)Manifest(Sub): يصف المادة التعليمية، حيث يصف ذاته ومادة تعليمية أخرى[3][4].



الشكل(1) ملف Manifest ضمن حزمة المحتوى

4. بناء محتوى تعليمي منسجم مع معيار SCORM

من أجل بناء محتوى تعليمي منسجم مع معيار SCORM كان لا بد أولاً من وضع مخطط للدرس المراد تطويره. لتكن لدينا البنية التالية للدرس بعنوان Logic Gates:

☒ **Logic Gates**

- Logic Gates Overview
- Main Gates
  - NOT gate (inverter)
  - AND gate
  - OR gate
- Sub-Gates
  - NAND gate
  - NOR gate
  - Negative-AND
  - Negative-OR
- Substituting one type of gate for another
  - Using NOR gates
    - Making a NOT gate from NOR gate
    - Making a AND gate from NOR gates
    - Making a NAND gate from NOR gates
    - Making a OR gate from NOR gates
  - Using NAND gates
    - Making a NOT gate from NAND gate
    - Making a AND gate from NAND gates
    - Making a OR gate from NAND gates
    - Making a NOR gate from NAND gates
- Example about logic Circuit
- Test

الآن، يمكننا إجراء مقابلة بين عناصر (مستويات) بنية الدرس السابقة وبين عناصر (مستويات) الحزمة في نموذج SCORM

☒ **Logic Gates (Organization)**

- Logic Gates Overview (Item) ➔ (resource) ➔ SCO01
- Main Gates (Item)
  - NOT gate (inverter) (Item) ➔ (resource) ➔ SCO02
  - AND gate (Item) ➔ (resource) ➔ SCO03
  - OR gate(Item) ➔ (resource) ➔ SCO04
- Sub-Gates(Item)
  - NAND gate(Item) ➔ (resource) ➔ SCO05
  - NOR gate(Item) ➔ (resource) ➔ SCO06
  - Negative-AND(Item) ➔ (resource) ➔ SCO07
  - Negative-OR(Item) ➔ (resource) ➔ SCO08
- Substituting one type of gate for another(Item)
  - Using NOR gates(Item)
    - Making a NOT gate from NOR gate(Item) ➔ (resource) ➔ SCO09
    - Making a AND gate from NOR gates(Item) ➔ (resource) ➔ SCO10
    - Making a NAND gate from NOR gates(Item)➔(resource)➔ SCO11
    - Making a OR gate from NOR gates(Item) ➔ (resource) ➔ SCO12
  - Using NAND gates(Item)
    - Making a NOT gate from NAND gate(Item) ➔ (resource) ➔SCO013
    - Making a AND gate from NAND gates(Item) ➔ (resource) ➔SCO14
    - Making a OR gate from NAND gates(Item) ➔ (resource) ➔ SCO15
    - Making a NOR gate from NAND gates(Item) ➔ (resource) ➔SCO16
- Example about logic Circuit (Item) ➔ (resource) ➔ SCO17
- Test(Item) ➔ (resource) ➔ SCO18

5. مرحلة بناء المحتوى التعليمي الخاضع لمعيار SCORM ضمن مستودع كائن التعلم:

تخزن عادة المواد التعليمية في قاعدة بيانات تدعى مستودع كائن التعلم(LOR). وقد أدى الاعتماد الواسع على معيار SCORM من قبل منظمات التعلم إلى نشوء عدد ضخم من المواد التعليمية الخاضعة لهذا المعيار وهذا أدى إلى مشاكل في إدارة تلك المواد. لذا، ركزت منظمة SCORM على ابتكار الطرق لحل تلك المشكلة وعلى إمكانية التعامل مع مستودعات كائنات التعلم(LORs) بحيث تلبي حاجات المستخدمين.

سعينا من خلال هذا البحث إلى وضع طريقة تساعد في بناء محتوى تعليمي ضمن مستودع كائن التعلم يلبي حاجة المستخدمين. وقد دعيت هذه الطريقة بمخطط إدارة المحتوى وفق نمط المستوي (LCMS) (Level-wise Content Management Scheme).

تتضمن مرحلة البناء هذه العمليات التالية[3]:

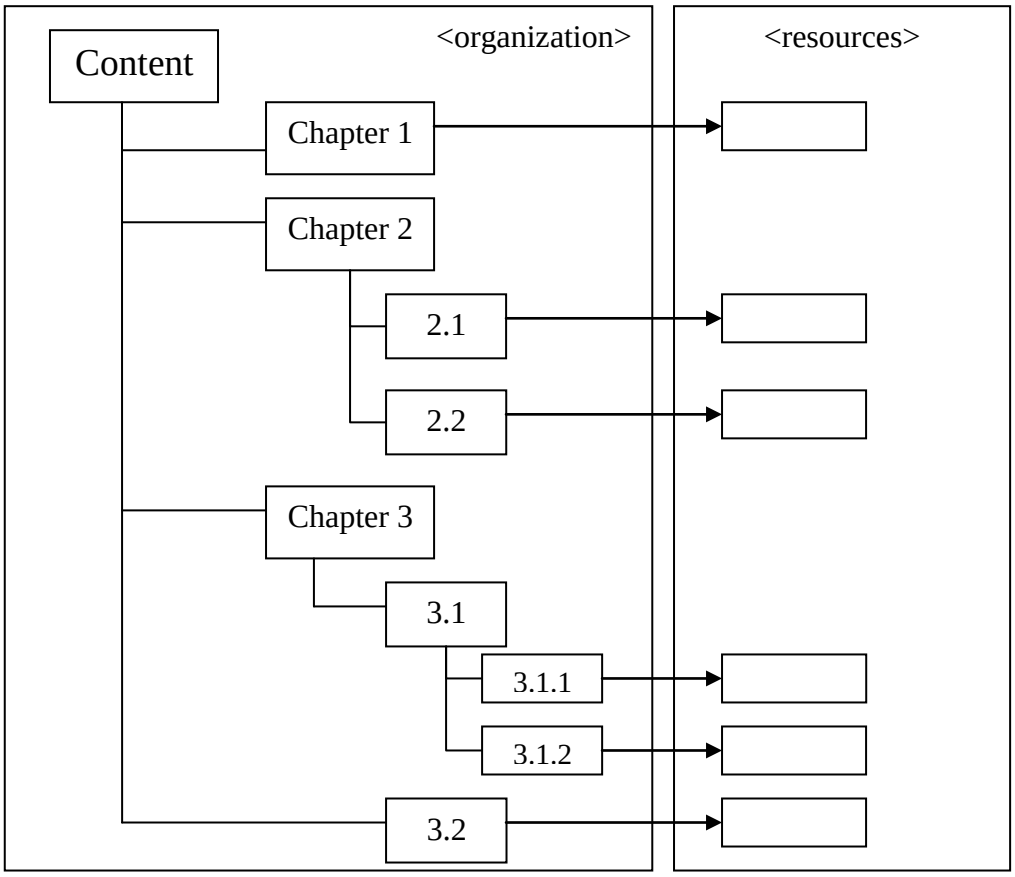
- عملية تحويل حزمة المحتوى إلى شجرة محتوى (CP2CT)(Content Package to Content Tree):

هذه العملية تنقل تركيب محتوى مواد التعلم الخاضعة لمعيار SCORM (حزمة المحتوى) إلى تركيب يشبه شجرة مع شعاع المميزة التمثيلي وبالعمق ذاته، يدعى شجرة نشاط (CT) وذلك لتمثيل كل مادة تعليمية.

- عملية التجميع العنقودي للمحتوى التعليمي وفق نمط المستوي (Level-wise Content Clustering Process): هذه العملية تجمع كائنات التعلم على شكل عناقيد وفقا لأشجار المحتوى (CTs) وتؤسس لمخطط التجميع العنقودي للمحتوى وفق نمط المستوي (LCCG) (level-wise content clustering graph) وتعيد بناء المخطط LCCG إذا كان ذلك ضرورياً.

المرحلة الأولى: عملية تحويل حزمة المحتوى إلى شجرة محتوى (CP2CT)

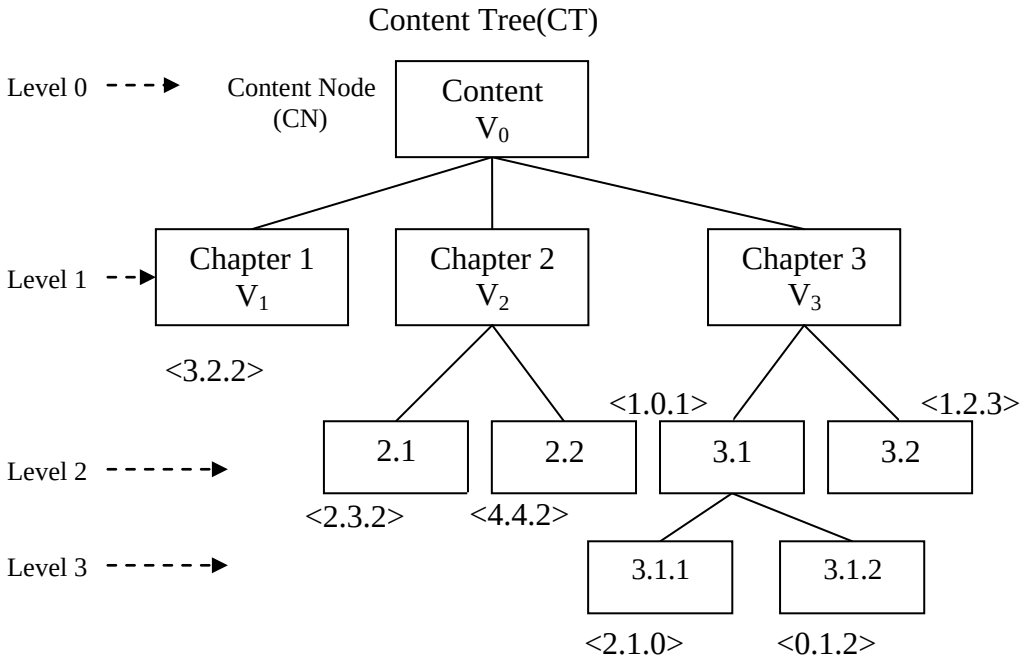
بفرض لدينا حزمة محتوى مؤلفة من ثلاثة فصول وفق الشكل (3).



الشكل (3): حزمة المحتوى

ومن أجل تحديد العلاقات بين كائنات التعلم (learning objects) (LOs) وفقاً لتركيب محتوى المواد التعليمية، تحول المعلومات المنظمة في حزمة المحتوى الخاضع لمعيار SCORM إلى تمثيل يشبه شجرة مع شعاع المميزة التمثيلي يدعى شجرة محتوى (Content Tree) (CT).

تأخذ شجرة المحتوى لحزمة المحتوى السابقة الشكل (4).



الشكل (4) شجرة المحتوى (CT) مع أشعة المميزة للعقد

في أي شجرة محتوى (CT) كل عقدة تدعى " عقدة محتوى (CN) (Content Node)" وترتبط ببعضها بعضاً بروابط (links)، ومنظمة تنظيماً هرمياً (Hierrarchical) لا يحتوى على حلقات (cycles) مغلقة.

ولتمثيل محتويات التعلم (مستند نصي) في أي عقدة محتوى(CN) تم اعتماد نموذج فضاء موجه (VSM) (Vector Space Model) والذي يعدّ النموذج الأكثر انتشارا في تعقب النص (إلى حد ما)، وقد اقترح من قبل Salton وآخرين عام 1975. في هذا النموذج مميزات المستند عبارة عن كلمات وتأتي قيم المميزة من مخططات الأوزان للتعابير المختلفة. يمكن توضيح هذا النموذج كالتالي[3][5] [6]:

كل مستند (عقد محتوى) يمثل بشعاع d في فضاء التعبير مثل $d = \{w_1, w_2, \ldots, w_n\}$ حيث wi: يمثل وزن التعبير i في المستند. ويأخذ i القيم i=1, 2, …,n. يحسب وزن التعبير من خلال تردد التعبير(Term Frequency) في المستند (wi = tfi) (عدد المرات التي يظهر فيها التعبير i في المستند). أما عامل تردد المستند العكسي IDF (Inverse Document Frequency) فيعطى بالعلاقة $\log(N/df_i)$ حيث تمثل N عدد المستندات الكلي وdf_i تمثل عدد المستندات التي يظهر فيها التعبير. من ثمة استنادا إلى مخطط الوزن TF×IDF (Term Frequency × Inverse Document Frequency) فإن وزن TF×IDF للتعبير يعطى بالعلاقة: $w_i = tf_i \times \log(N/df_i)$. أي أن التعابير التي تتكرر بشكل أكبر في مستند معين ولكن تكرارها أقل في مستندات أخرى تعطى أوزاناً أعلى في ذلك المستند[5][6].

ولتمثيل المستند تم استخدام مخطط دليل المستند DIG(Document Index Graph). حيث يفهرس هذا النموذج المستندات مع المحافظة على تركيب الجملة في المستندات الأصلية.

يعرف البيان لهذا المخطط على النحو التالي:

$$G = (V , E)$$

حيث V: مجموعة العقد {v1, v2, … , vn} وحيث تمثل كل عقدة v كلمة مفردة في مجموعة المستندات الكلية.

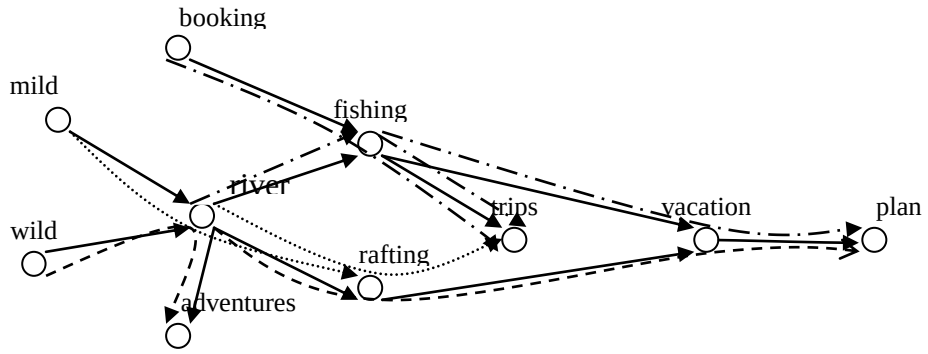
E: مجموعة الحواف (edges) {e1, e2, … , em} حيث تمثل كل حافة e زوجاً منظماً من العقد (v_i, v_j). حيث أن الحافة (v_i , v_j) هي من العقدة vi إلى العقدة vj إذا وفقط إذا ظهرت الكلمة v_j على نحو متعاقب للكلمة v_i في أي مستند.

وفقاً لهذا التعريف يكون عدد العقد في المخطط هو عدد الكلمات الفريدة في مجموعة المستندات.

تحمل العقد في المخطط المعلومات عن المستندات التي تظهر فيها ومعلومات عن مسار الجملة.

فإذا كان لدينا جملة مؤلفة من m كلمة تظهر في مستند واحد يتألف من سلسلة الكلمات التالية: {v_1, v_2, … , v_m}. تمثل الجملة في المخطط من خلال المسار من العقدة v_1 إلى العقدة v_2، وتمثل (v_1 , v_2) و (v_2 , v_3) و (v_3 , v_4) و……(v_{m-1} , v_m) الحواف. وتخزن معلومات المسار في القمم على طول المسار لتمييز كل جملة منفردة. إن الجمل التي تشترك بالعبارات الثانوية ستملك أجزاءً مشاركة لمساراتها في المخطط والذي يقابل العبارة الثانوية المشتركة[5].

فإذا فرضنا وجود ثلاثة مستندات وكل مستند يحوي على عدد من العبارات مع بعض التوافق بين المستندات. تمثل الجملة من المستند الأول بالسهم ▸..... والجملة من المستند الثاني بالسهم ▸ – – – والجملة من المستند الثالث بالسهم ▸– · – كما في الشكل (5).



Document 1	Document 2	Document 3
river rafting mild river rafting river rafting trips ▸	Wild river adventures River rafting vacation plan – – – ▸	Fishing trips Fishing vacation plan Booking fishing trips River fishing — · — ▸

الشكل (5) مخطط دليل المستند DIG

فإذا كانت لدينا المفاهيم التمثيلية التالية: {mild, vacation, trips} ومعبر عنها وفق الجدول (1).

الجدول(1)

	عدد مرات ظهور الكلمة mild	عدد مرات ظهور الكلمة vacation	عدد مرات ظهور الكلمة trips		عدد المستندات الكلي N	عدد المستندات التي يظهر فيها التعبير df_i	عامل تردد المستند العكسي $Log(N/df_i)$
المستند الأول	$W_1 = tf_1 = 1$	$W_2 = tf_2 = 0$	$W_3 = tf_3 = 1$		3	df_1=1	0.477
المستند الثاني	$W_1 = tf_1 = 0$	$W_2 = tf_2 = 1$	$W_3 = tf_3 = 0$			df_2=2	0.176
المستند الثالث	$W_1 = tf_1 = 0$	$W_2 = tf_2 = 1$	$W_3 = tf_3 = 2$		3	df_3=2	0.176

عندها يكون شعاع المميزة باستخدام قانون TF × IDF بالنسبة للمستند الأول:

$$\text{Document 1} = \langle tf_1 \times idf_1, tf_2 \times idf_2, \ldots, tf_n \times idf_n \rangle$$
$$= \langle tf_1 \times \log(N/df_1), tf_2 \times \log(N/df_2), tf_3 \times \log(N/df_3) \rangle$$
$$\langle 1\times 0.477, 0\times 0.176, 1\times 0.176 \rangle = \langle 0.477, 0, 0.176 \rangle$$

بالمثل نحصل على أشعة المميزة لبقية المستندات.

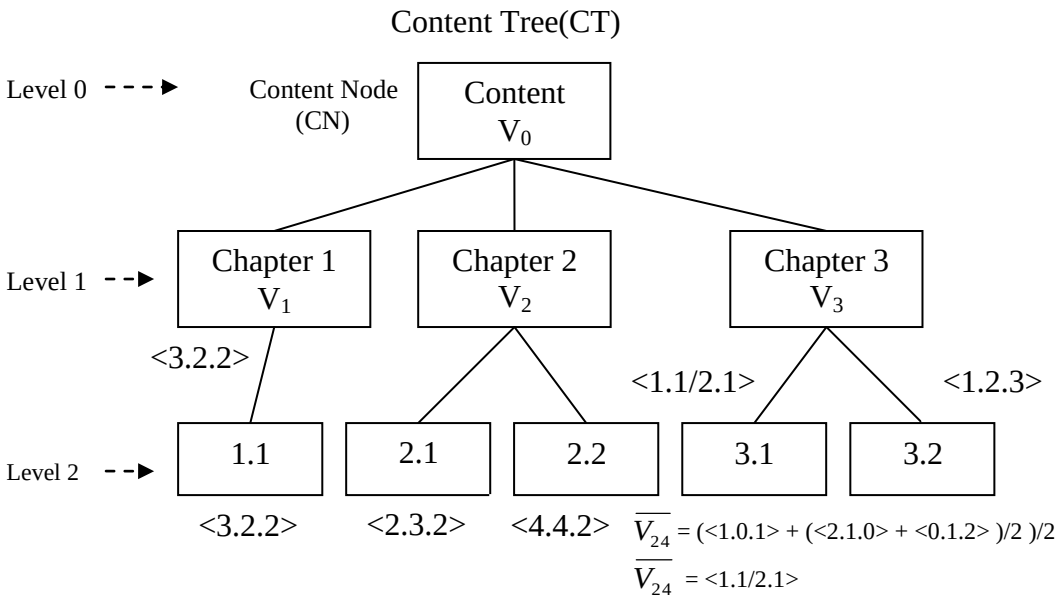
بتعميم ذلك على عقد المحتوى لأشجار النشاط نحصل على شجرة المحتوى ذات أشعة مميزة لكل عقدة ممثلة محتويات التعلم وهذا موضح في الشكل (5).

في معظم الأحيان عمق تركيب المحتوى متنوع لذا فإنه لجعل عملية التجميع العنقودية فعالة يتم جعل جميع أشجار النشاط ذات عمق واحد. فإذا كان عمق ورقة ما في شجرة محتوى قصيراً جداً عندها يتم إضافة عقدة افتراضية (Virtual Node) (VN) على نحو متكرر كعقدة فرعية (child node) حتى يزول الاختلاف في العمق، وسيكون شعاع المميزة لكل عقدة افتراضية مماثلاً لشعاع المميزة الخاص بعقدة المحتوى الأصلية (parent CN) أو للعقدة الافتراضية الناشئة عنها. أما إذا كان عمق الورقة كبيراً جداً عندها سيتم دمج المعلومات الخاصة بالعقد الفرعية في عقدة محتوى جديدة وستملك هذه العقدة شعاع مميزة يمثل معدل(Average) الأشعة المميزة للعقد الفرعية المشكلة للعقدة الجديدة.

بتطبيق ذلك على المحتوى المفروض الموضح في الشكل (2) نجد أن عمق العقدة V_1 "Chapter 1" ليس كبيراً عندها يتم إضافة عقدة افتراضية تدعى V_{11} يكون لها شعاع المميزة $\overline{V_{11}}$ = <3.2.2> وهو مماثل لشعاع المميزة $\overline{V_1}$. كما أن عقدة المحتوى $V_{3.1}$ طويلة جداً وتتضمن عقداً فرعية وهي $V_{3.1.1}$ و $V_{3.1.2}$ لذا يتم دمجهما في عقدة محتوى واحدة وهي $V_{3.1}$ ويكون شعاع المميزة الخاص بهذه العقدة الجديدة $\overline{V_{24}}$ عبارة عن معدل كل من <1.0.1> و 2/(<2.1.0> + <0.1.2>) أي

$$\overline{V_{24}} = (\langle 1.0.1 \rangle + (\langle 2.1.0 \rangle + \langle 0.1.2 \rangle)/2)/2 = \langle 1.1/2.1 \rangle$$

تأخذ شجرة المحتوى بعد إجراء عملية CP2CT الشكل (6).



الشكل (6) منظومة المحتوى المحولة إلى شجرة محتوى مع أشعة المميزة

خوارزمية عملية تحويل حزمة المحتوى إلى شجرة محتوى(CP2CT):

1. تعريف الرموز:

CP: يشير إلى حزمة المحتوى المشكلة وفق معيار SCORM.

CT: يشيرإلى شجرة المحتوى المحولة من حزمة محتوى.

CN: يشير إلى عقدة المحتوى في شجرة المحتوى.

CN_{leaf}: يشير إلى عقدة الورقة التابعة لعقدة محتوى في شجرة المحتوى.

DCT: يشير إلى عمق شجرة المحتوى المرغوب.

DCN: يشير إلى عمق شجرة المحتوى.

Input: عبارة عن حزمة المحتوى المشكلة وفق معيار SCORM.

Output: عبارة عن شجرة المحتوى مع شعاع المميزة.

2. خطوات الخوارزمية:

الخطوة 1: من أجل كل مادة <item> في حزمة المحتوى نقوم بما يلي:

1.1 ننشئ عقدة محتوى مع شعاع مميزة وذلك استناداً إلى مخطط الموازنة TF×IDF.

2.1 ثم نقوم بإدخاله في المستوي المقابل في شجرة المحتوى.

الخطوة 2: من أجل كل ورقة تابعة لعقدة محتوى(CN_{leaf}) في شجرة المحتوى نقوم بما يلي:

إذا كان عمق ورقة عقدة المحتوى (CN_{leaf}) < العمق المرغوب لشجرة المحتوى(DCT)

عندها سيتم إدخال عقدة افتراضية (VN) على نحو متكرر كعقدة فرعية منها حتى يصبح عمق ورقة عقدة المحتوى = العمق المرغوب لشجرة المحتوى(DCT)

وإلا: إذا كان عمق ورقة عقدة المحتوى(CN_{leaf}) > العمق المرغوب لشجرة المحتوى(DCT)

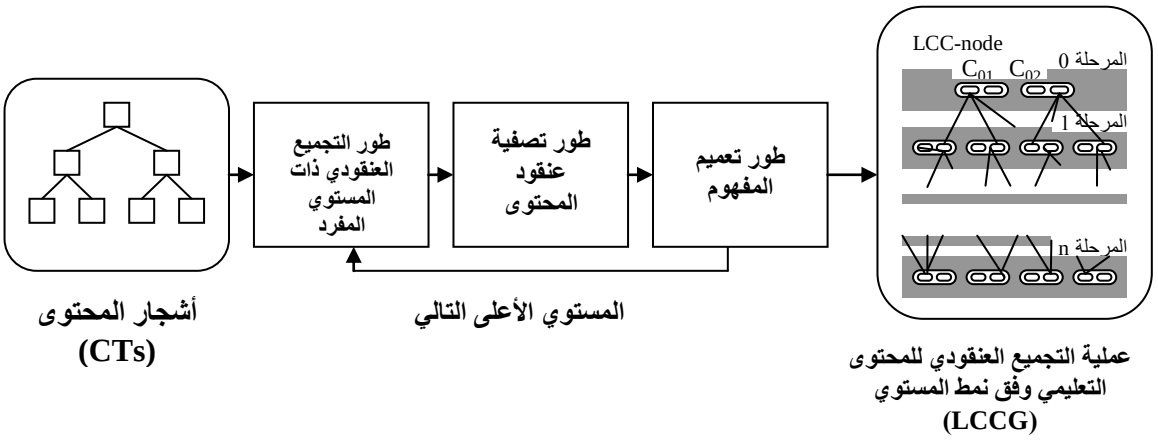
عندها ستدمج عقدة المحتوى الأصلية للمعلومات الخاصة بالعقد الفرعية التابعة، وتقوم بعملية السحب للأعلى (rolling up) وذلك لحساب معدل الأشعة المميزة لكل عقدة فرعية بحيث يصبح عمق العقدة الأصلية مماثلاً للعمق المرغوب لشجرة المحتوى(DCT)

الخطوة 3: يتم الحصول على شجرة المحتوى (CT) مع شعاع المميزة.

المرحلة الثانية: عملية التجميع العنقودي للمحتوى التعليمي وفق نمط المستوي (LCCG)

بعد أن يتم تحويل المعلومات المنظمة لحزمة المحتوى إلى شجرة محتوى (CT)، يمكن تطبيق التقنية العنقودية لتحديد العلاقات بين عقد المحتوى(CNs) في شجرة المحتوى CT.

وتم تقسيم العملية إلى ثلاثة أطوار وهي: 1) طور التجميع العنقودي ذو المستوي المفرد، 2) طور تصفية عنقود المحتوى، 3) طور تعميم المفهوم. وهذا موضح من خلال الشكل (7).



الشكل (7) مخطط تدفق خوارزمية التجميع العنقودي للمحتوى وفق نمط المستوي

1) طور التجميع العنقودي ذو المستوي المفرد (Single Level Clustering Phase):

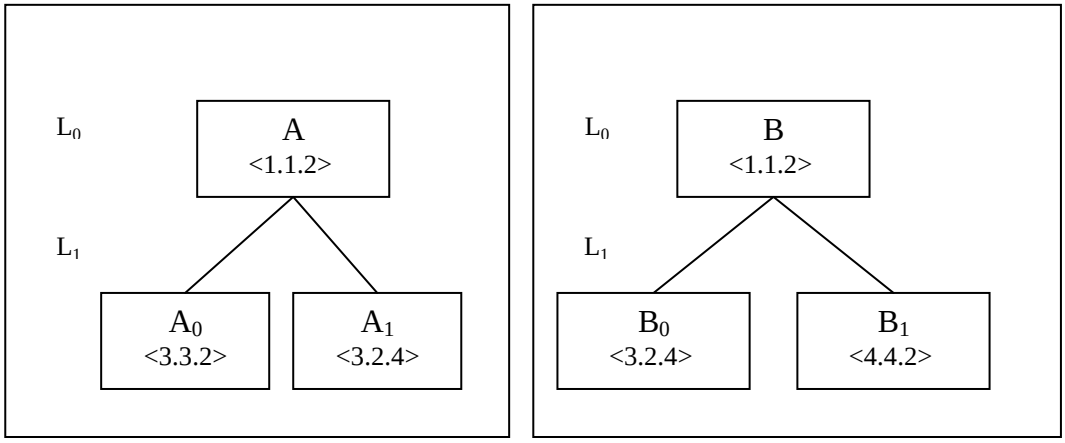
في هذا الطور تجمع عقد المحتوى (CNs) التي تنتمي للمستوى ذاته من أشجار المحتوى(CTs). وفق عتبة تشابه معينة. تبدأ عملية التجميع العنقودي للمحتوى في المستوى الأدنى وتتقدم باتجاه المستوى الأعلى في شجرة المحتوى، ثم تخزن نتائج التجميع في مخطط LCCG. ويتم تجميع عقد المحتوى من خلال قياس التشابه (similarity measure) المستخدم على نطاق واسع في مجال التجميع العنقودي للمستندات النصية[5][6][7] .

$$\text{Similarity} = \text{cosine}(V_A, V_B) = \frac{V_A . V_B}{|V_A|\,|V_B|}$$

حيث V_A و V_B عبارة عن أشعة المميزة للعقدتين N_A وN_B على التوالي.

لتوضيح ذلك نفرض وجود شجرتي محتوى وهما CT_A و CT_B حيث تنتمي العقدتان N_{A0} وN_{A1} إلى العقدة N_A، وتنتمي العقدتان N_{B0} وN_{B1} إلى العقدة N_B.

يوضح الشكل (6) شجرتي المحتوى N_A وN_B مع أشعة المميزة.



الشكل (8) شجرتا المحتوى CT_A و CT_B

من خلال الشكل (8) نجد أنه لدينا مستويان L_0 و L_1. تبدأ عملية التجميع وفق المستوي الأدنى L_1 باتجاه المستوي الأعلى L_0. يتم قياس التشابه بين كل عقدتين في المستوي الأدنى الذي يحتوي على أربع عقد محتوى وفي A_0 و A_1 وB_0 وB_1 كما يلي:

- قياس التشابه بين العقدتين A_0 وB_0 [5][6][7]:

$$\text{Similarity} = \text{cosine}(V_{A0}, V_{B0}) = \frac{V_{A0}.V_{B0}}{|V_{A0}|\,|V_{B0}|}$$

$$= \frac{3*3+3*2+2*4}{\sqrt{(3^2+3^2+2^2)(3^2+2^2+4^2)}} = \frac{23}{25.2259} = 0.91 < T$$

حيث T=1 وهي تمثل عتبة التشابه.

أي أن العقدتين V_{A0} و V_{B0} لا ينتميان إلى العنقود ذاته.

حيث تنتمي العقدة V_{A0} إلى العنقود C_{11} والعقدة V_{B0} إلى العنقود C_{12}.

- قياس التشابه بين العقدتين A_0 وB_1

$$\text{Similarity} = \text{cosine}(V_{A0}, V_{B1}) = \frac{V_{A0}.V_{B1}}{|V_{A0}|\,|V_{B1}|}$$

$$= \frac{3*4+3*4+2*2}{\sqrt{(3^2+3^2+2^2)(4^2+4^2+2^2)}} = \frac{28}{28.1425} = 0.995 < T$$

أي أن العقدتين V_{A0} و V_{B1} لا تنتميان إلى العنقود ذاته. إن العقدة V_{B1} تنتمي إلى العنقود C_{13}.

- قياس التشابه بين العقدتين A_1 وB_0

$$\text{Similarity} = \text{cosine}(V_{A1}, V_{B0}) = \frac{V_{A1}.V_{B0}}{|V_{A1}|\,|V_{B0}|}$$

$$= \frac{3*3+2*2+4*4}{\sqrt{(3^2+2^2+4^2)(3^2+2^2+4^2)}} = \frac{29}{29} = 1 = T$$

أي أن العقدتين A_1 و B_0 تنتميان إلى العنقود ذاته، وبما أن العقدة A_1 تنتمي إلى العنقود C_{12} فإن العقدة B_0 تنتمي إلى العنقود C_{12} أيضاً.

- قياس التشابه بين العقدتين A_1 و B_1

$$\text{Similarity} = \text{cosine}(V_{A1}, V_{B1}) = \frac{V_{A1}.V_{B1}}{|V_{A1}|\,|V_{B1}|}$$

$$= \frac{3*4+2*4+4*2}{\sqrt{(3^2+2^2+4^2)(4^2+4^2+2^2)}} = \frac{28}{32.311} = 0.867 < T$$

أي أن العقدتينV_{A1} و V_{B1} لا تنتميان إلى العنقود ذاته.

الآن يتم الانتقال للمستوي الأعلى L_0 وإجراء قياس التشابه في هذا المستوي.

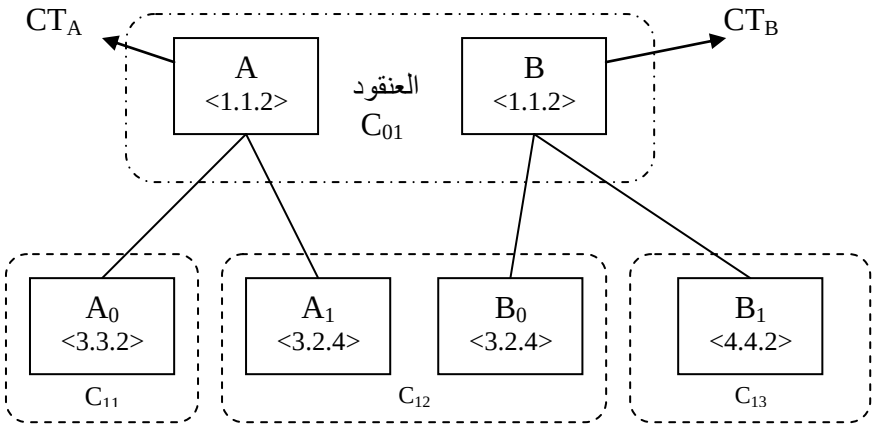
- قياس التشابه بين العقدتين A و B

$$\text{Similarity} = \text{cosine}(V_A, V_B) = \frac{V_A.V_B}{|V_A|\,|V_B|}$$

$$= \frac{1*1+1*1+2*2}{\sqrt{(1^2+1^2+2^2)(1^2+1^2+2^2)}} = \frac{6}{6} = 1 = T$$

أي أن العقدتين V_A و V_B تنتميان إلى العنقود ذاته وهو العنقود C_{01}.

تصبح شجرتا المحتوى بعد إجراء عملية التجميع العنقودي وفق نمط المستوي وفق الشكل (9).



الشكل (9) عملية التجميع العنقودي للمحتوى وفق المستوي المفرد

لتسهيل التجميع العنقودي للمحتوى يجب أن يكون عدد مراحل عملية التجميع العنقودي مساوياً لعمق شجرة المحتوى، وتعالج كل مرحلة نتيجة التجميع العنقودي لعقد المحتوى في المستويات المقابلة لأشجار المحتوى. حيث تخزن المميزة العنقودية (Cluster Feature) (CF) المعلومات المرتبطة بالعنقود. وهذه المميزة معرفة كالتالي[3]:

$$CF = (N, \overline{VS}, CS)$$

حيث N: عدد عقد المحتوى (CNs) في العنقود.

$\overline{VS} = \sum_{i=1}^{N} \overline{V_i}$: مجموع أشعة المميزة $\overline{V}$ لعقد المحتوى.

$CS = \left| \sum_{i=1}^{N} \overline{V_i} / N \right| = \left| \overline{VS} / N \right|$: قيمة معدل مجموع أشعة المميزة في العنقود. تمثل $(\overline{VS} / N)$ مركز العنقود(CC) (Cluster Center).

إذا كان لدينا عنقود ما C_0 مخزن فيه العقدة N_A تحوي أربع عقد محتوى، ومن ثمة وبالتالي تتضمن أربع موجهات مميزة وهي: <3.3.2> و <3.2.2> و <2.3.2> و <4.4.2> عندها يكون

$$\overline{VS} = <3.3.2> + <3.2.2> + <2.3.2> + <4.4.2> = <12.12.8>$$

$$\overline{VS}/4 = <12.12.8>/4 = <3.3.2>$$

$$CS = \sqrt{9+9+4} = 4.69$$

ومن ثمة يكون $CF_A = (N, \overline{VS}, CS) = (4, <12.12.8>, 4.69)$

فإذا تم إضافة عقدة جديدة ذات شعاع مميزة <8.3.2> إلى العنقود C_0 يصبح CF_A كما يلي:

$$CF_A = (N, \overline{VS}, CS) = (5, <20.15.10>, 5.385)$$

خوارزمية التجميع العنقودي وفق المستوي المفرد:

- **تعريف الرموز:**

CNset: مجموعة عقد المحتوى (CNs) في المستوى ذاته (L) من أشجار المحتوى (CTs).

T: عتبة التشابه من أجل عملية التجميع.

Input: عبارة عن CNset و T .

Output: عبارة عن مجموعة العقد الناتجة عن تجميع المحتوى وفق نمط المستوى(LCC-Nodes)، الذي يخزّن نتائج التجميع لأشجار المحتوى.

- **خطوات الخوارزمية:**

الخطوة 1: نقوم بإدخال عقدة المحتوى $n_i \in CN_{set}$ في عنقود ضمن LCC-Node.

الخطوة 2: $\forall n_i \in CN_{set}$.

1.2 إذا وجد عنقود ما بقيمة تشابه > عتبة التشابه (T).

عندها ندخل هذه العقدة n_i في هذا العنقود ومن ثم نقوم بتحديث مميزة العنقود (CF) في LCC-Node.

وإلا ندخل العقدة n_i في عنقود جديد مخزن في LCC-Node جديدة.

الخطوة 3: إرجاع مجموعة العقد الناتجة عن تجميع المحتوى وفق نمط المستوي(LCC-Nodes).

2) طور تصفية عنقود المحتوى (Content Cluster Refining Phase):

بهدف تعديل دقة العناقيد يحسب التشابه بين عنقودين بقياس التشابه بين مركزي كل عنقودين وفق القانون التالي:

$$Similarity = \cos(CC_A.CC_B) = \frac{CC_A.CC_B}{|CC_A||CC_B|} = \frac{(\overline{VS}/N_A).(\overline{VS}/N_A)}{CS_A.CS_B}$$

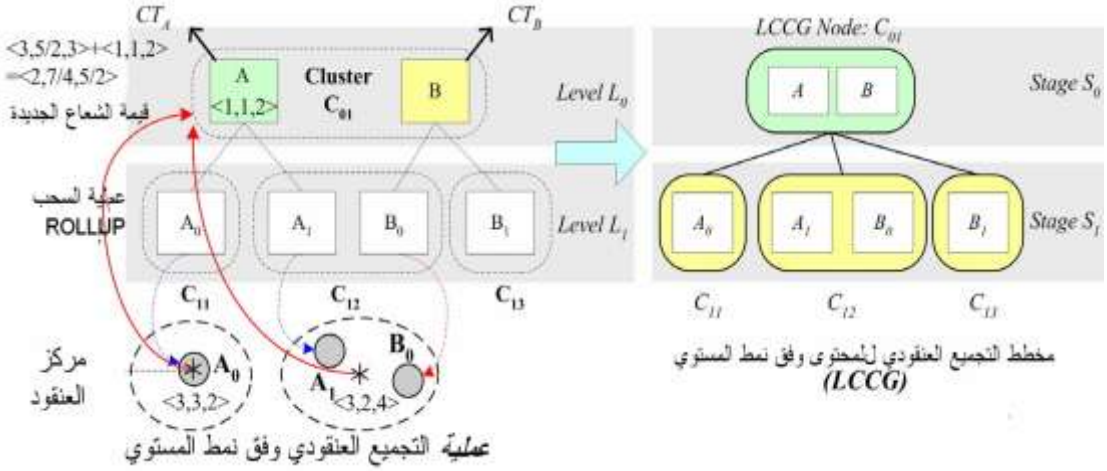
وبعد حساب التشابه إذا كان هناك عنقودان ينبغي دمجهما في عنقود جديد، عندها تكون مميزة العنقود (CF) لهذا العنقود الجديد هي:

$$CF_{new} = (N_A + N_B, \overline{VS_A} + \overline{VS_B}, \left|(\overline{VS_A} + \overline{VS_B})/(N_A + N_B)\right|).$$

3) طور تعميم المفهوم (Concept Generalizing Phase):

إن مجمل عملية تجميع المحتوى وفق نمط المستوي تتمثل وفق الشكل (10).

من الشكل (10) نجد أن عدد مراحل عملية التجميع (S_0, S_1) مساو لعمق شجرة المحتوى (L_0, L_1).



الشكل (10) مثال يوضح مخطط تجميع المحتوى وفق نمط المستوي

الخوارزمية الكلية للتجميع العنقودي للمحتوى وفق نمط المستوي:

- **تعريف الرموز:**

D: عمق شجرة المحتوى (CT).

L_0 ~ L_{D-1}: يشير إلى مستويات شجرة المحتوى، بدءاً من المستوى الأعلى نزولاً إلى المستوي الأدنى.

S_0 ~ S_{D-1}: يشير إلى مراحل خوارزمية تجميع المحتوى وفق المستوي(LCCG).

T_0 ~ T_{D-1}: تشير إلى عتبات التشابه لتجميع عقد المحتوى (CNs) في المستويات L_0 ~ L_{D-1} على التعاقب.

CTset: مجموعة شجرات المحتوى (CTs) ذات العمق الواحد(D).

CNset: عقد المحتوى(CNs) التي تنتمي لمستوي الشجرة ذاته(L).

Input: CTset.

Output: خوارزمية تجميع المحتوى وفق المستوي والتي تحجز(holds) نتائج التجمع في كل مستوي شجرة المحتوى.

- **خطوات الخوارزمية:**

الخطوة 1: من أجل i = L_{D-1} إلى L_0 قم بإجراء الخطوات 2 وحتى 4.

الخطوة 2: تجميع المستوي المفرد(Single Level Clustering):

2.1 CNset = عقد المحتوى Ǝ CTset في L_i.

2.2 تشغيل خوارزمية تجميع المستوي المفرد (SLCAlg) من أجل CNset ذات العتبة T_i.

الخطوة 3: تصفية عنقود المحتوى:

3.1 نفذ الخطوات الثانوية التالية(3.2– 3.4) على نحو متكرر حتى ينعدم الاختلاف بين تكرارين.

3.2 CNset = العقد في مركز العنقود (CC). مجموع العقد الخاصة بتجميع المحتوى وفق المستوي(LCC-Nodes) في S_i.

3.3 تشغيل خوارزمية تجميع المستوي المفرد(SLCAlg) من أجل CNset بعتبة T_i.

3.4 تخزين العناقيد الناتجة في العقد الخاصة بتجميع المحتوى وفق المستوي من خوارزمية تجميع المحتوى وفق المستوي(LCCG) في المرحلة S_i.

الخطوة 4: تعميم المفهوم:

4.1 إذا كان $L_0 \neq i$

عندها نقوم بإنجاز عملية السحب للأعلى (roll-up) وذلك لحساب أشعة المميزة لعقد المحتوى من المستوي L_{i-1}.

الخطوة 5: ناتج مخطط التجميع العنقودي للمحتوى وفق نمط المستوي.

النتائج والمناقشة:

حصلنا من خلال الدراسة على النتائج التالية:

1. بناء وتحزيم محتوى تعليمي وفق معيار SCORM.
2. ساعد مفهوم فضاء الحالة وشعاع المميزة في توحيد أعماق شجرات المحتوى.
3. تم استخدام مخطط دليل المستند DIG(Document Index Graph) لتمثيل الدرس وتم استخدام المعادلات الرياضية في الحصول على المعلومات التي تعبر عن هذا الدرس من خلال الحصول على شعاع المميزة للدرس المفترض.
4. تم تجميع عقد المحتوى في عناقيد من خلال قياس التشابه بين كل عقدين من عقد المحتوى للأشجار CTs باعتماد عتبة التشابه T=1 وتخزين معلومات العناقيد من خلال حساب المميزة العنقودية (CF) لكل عنقود.
5. تم تعديل دقة العناقيد من خلال حساب التشابه بين عنقودين.
6. تم تعميم عملية التجميع العنقودي وفق كل مستوي بدءاً من المستوي الأدنى وانتهاءً بالمستوي الأعلى.

الاستنتاجات:

أمكن التغلب على المشاكل الناجمة عن كثرة وتنوع الموارد التعليمية المشكلة وفق معيار SCORM والذي ينجم عنه سوء إدارة تلك المواد من خلال التركيز على المعلومات االمعبّرة عنها، والتي تم التوصل إليها من خلال تمثيل المحتوى التعليمي بشعاع مميزة وتنظيم محتويات التعلم وفقاً لأشعة المميزة هذه، وذلك استنادا إلى مجموعة من الخوارزميات التي تنظم عملية تجميع المحتوى وتخزينه ضمن مستودع كائن التعلم.

كما تم الاستعاضة عن تخزين المحتوى التعليمي في المستودعات بتخزين معلومات المراكز العنقودية، والتي تم التوصل إليها من خلال تحويل حزمة المحتوى إلى شجرة محتوى، وتمثيل كل عقدة محتوى بشعاع مميزة يعبر عن معلومات هذه العقدة، ومن ثم تجميع عقد المحتوى في عناقيد باعتماد قانون قياس التشابه(similarity measure)

بين كل عقدتين من عقد أشجار المحتوى ووفقاً لعتبة تشابه محددة. وتم تعديل دقة العناقيد من خلال قياس التشابه للعناقيد الناتجة.

المراجع:

1. STRATKIS, M.; CHRISTOPHIDES, V. *WP2 Deliverable 2.1:E-Learning Standards*. 2003. January 06, 2007. <http://www.dcs.bbk.ac.uk/selene/reports/Del21.pdf>
2. OSTYN, C. *In the Eye of the SCORM, An introduction to SCORM 2004 for* Content Developer. 2005. April 21, 2007. <http://www.ostyn.com/resscormtech.htm>
3. JUN-MING SU; SHIAN-SHYONG TSENG; CHING-YAO WANG; YING-CHIEH LEI; YU-CHANG SUNG; WEN-NUNG TSAI. *A Content Management Scheme in a SCORM Compliant Learning Object Repository*, Journal of information science and engineering 21, 2005, 1053-1075. June 04, 2007. <http://www.iis.sinica.edu.tw/JISE/2005/200509_13.pdf>
4. *IMS Content Packaging Information Model*, 04 October 2004. August 04, 2007. <http://www.imsglobal.org/content/packaging/cpv1p1p4/imscp_infov1p1p4.htm>
5. HAMMOUDA, K.; KAMEL M. *Data Mining in e-Learning*. July 26, 2007. <http://pami.uwaterloo.ca/pub/hammouda/hammouda-learning.pdf>
6. KHEIRBEK, A. *Information Retrieval*.University of Damascus Faculty of Information technology, 2004, 142.
7. ABOUD, M. *Information Retrieval*. Damascus University Faculty of Information Technology, 2004, 42.