



UNIUNEA EUROPEANĂ
Fondul Social European



GUVERNUL ROMÂNIEI
Ministerul Dezvoltării Regionale
și Administrației Publice



Inovație în administrație
Programul Operațional
"Dezvoltarea Capacității
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013

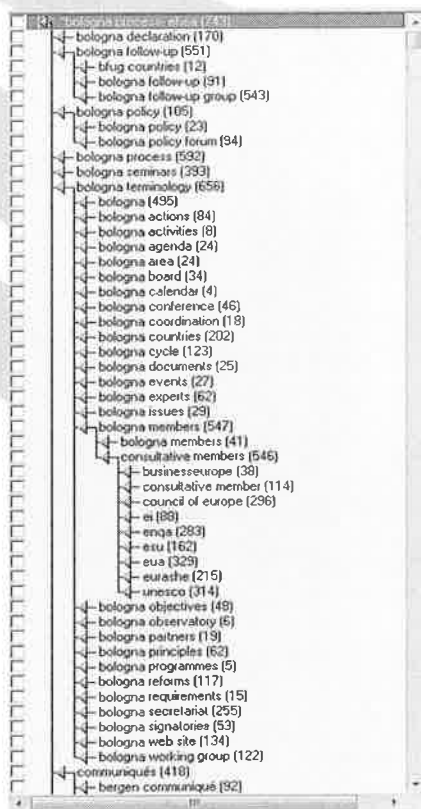
RAPORT PRIVIND SETUL DE INSTRUMENTE SEMANTICE PENTRU ANALIZA AUTOMATĂ A DISCURSULUI BOLOGNA

Realizarea, în cadrul proiectului, a patru analize semantice cu privire la Procesul Bologna –EHEA, presupune, dincolo de elaborarea unei metodologii adecvate, realizarea prealabilă a unui set de instrumente semantice adecvate.

Trebuie subliniat de la început că aceste instrumente sunt menite să funcționeze integrate în programele: *Tropes*, *Zoom/Owledge*, atât în limba engleză, cât și în română. Dacă prima condiție se referă la adecvarea informatică, a doua are în vedere adecvarea informațională sau lexicografică. Le vom prezenta pe rând.

PERSPECTIVA INFORMATICĂ

Atât *Tropes*, cât și *Zoom/Owledge* folosesc, în analizele lor, dicționare proprii numite **scenarii** (thesauri) și care sunt create sau adaptate de specialist în funcție de obiectivele specifice fiecărei analize. Este vorba de fișiere de tip *.scn (proprietary) și care pot fi vizualizate arborescent:



Scenariul „Bologna” în *Tropes* sau *Zoom* (vedere parțială)



UNIUNEA EUROPEANĂ
Fondul Social European



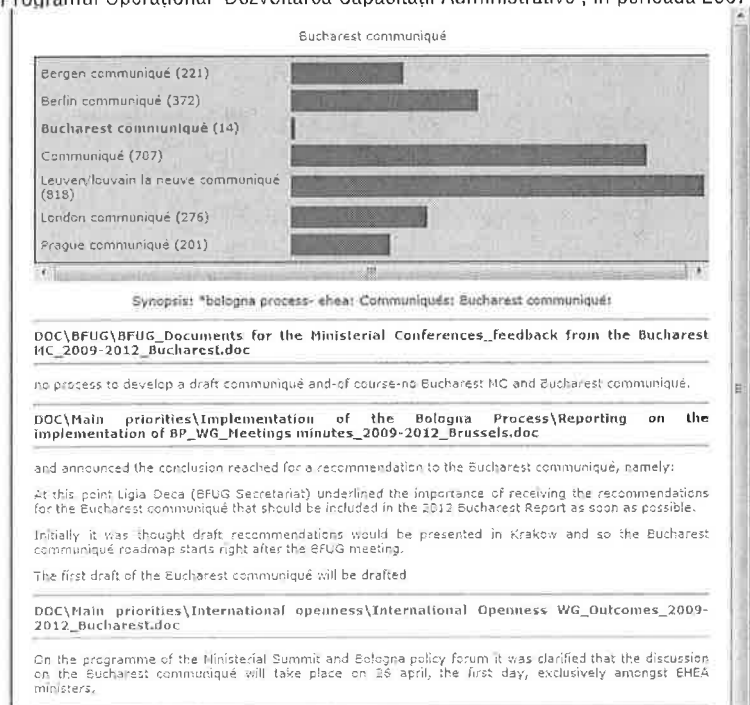
GUVERNUL ROMÂNIEI
Ministerul Dezvoltării Regionale
și Administrației Publice



Inovație în administrație
Programul Operațional
"Dezvoltarea Capacității
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013

- Show documents
Show graph (Synopsis)
Choose:
- education (276952)
 - *bologna process- ehea (100431)
 - bologna declaration (853)
 - bologna follow-up (8966)
 - bfug countries (16)
 - bologna follow-up (213)
 - bologna follow-up group (8731)
 - bologna policy (864)
 - bologna policy (33)
 - bologna policy forum (931)
 - bologna process (7148)
 - bologna seminars (3478)
 - bologna terminology (19820)
 - communiqués (2605)
 - bergen communiqué (221)
 - berlin communiqué (372)
 - bucharest communiqué (14)
 - communiqué (707)
 - leuven/louvain la neuve communiqué (818)
 - london communiqué (276)
 - prague communiqué (201)
 - ehea (9798)
 - main priorities (45672)
 - ministerial conferences (1091)
 - bucharest ministerial conference (45)
 - budapest vienna ministerial conference (3)
 - ministerial conference (1043)
 - networks (111)
 - sarbonne declaration (83)
 - geographical entities (41295)
 - society (34954)
 - research and technology (34789)
 - health (26111)
 - economy (22833)
 - politics (18518)
 - communications (15877)
 - legal terms (9570)
 - proper names (9515)
 - cultura (5907)
 - science(s) (5265)
 - european union (3584)



Scenariul „Bologna” pe Web (vedere parțială)

Tropes și Zoom permit exportul întregului scenariu via Web pentru ca analiza să poată fi vizualizată de întreaga comunitate căreia i se adresează.

Pentru crearea unui astfel de scenariu, autorii programului Tropes (software-ul folosit ca Scenario builder) prezintă în manualul de referință următoarele indicații:

„Tropes is supplied with a number of default Scenarios, rich in classifications, which can serve as a starting point for the rapid creation of a new classification plan. We advise you to be careful to keep the Scenarios supplied with the software, and to save them under a new name if you want to modify them.

A Scenario consists of a number of Semantic Groups, i.e. groups of words and/or Equivalent classes, which can be arranged in a hierarchy of nine ranks. Most combinations can be made by means of the mouse, either in the Scenario tool or interactively with Tropes' main window. Just create a group and put something into it.

When viewing a Reference in the Result frame, you can add it directly to the Scenario by dragging it with the mouse and dropping it on the Scenario tool. In Tropes Version 6 and later, you can also add the verb and adjectives lemmas (or their lemmas in languages other than English) directly in the Scenario. It is also possible to grab a word in Tropes [Text] frame and drag it onto the Scenario. Use the [Ctrl] and [Shift] keys to control the addition or the insertion during this operation.

The Result popup menu (right click with the mouse) also contains two functions linked to the Scenario. The command [Add to the Scenario] creates an entry having the same group name as the



UNITATEA EXECUTIVĂ PENTRU
FINANȚAREA ÎNVĂȚĂMÂNTULUI
SUPERIOR, A CERCETĂRII,
DEZVOLTĂRII ȘI INOVĂRII



Politici publice fundamentate
în Învățământul Superior



UNIUNEA EUROPEANĂ
Fondul Social European



GUVERNUL ROMÂNIEI
Ministerul Dezvoltării Regionale
și Administrației Publice

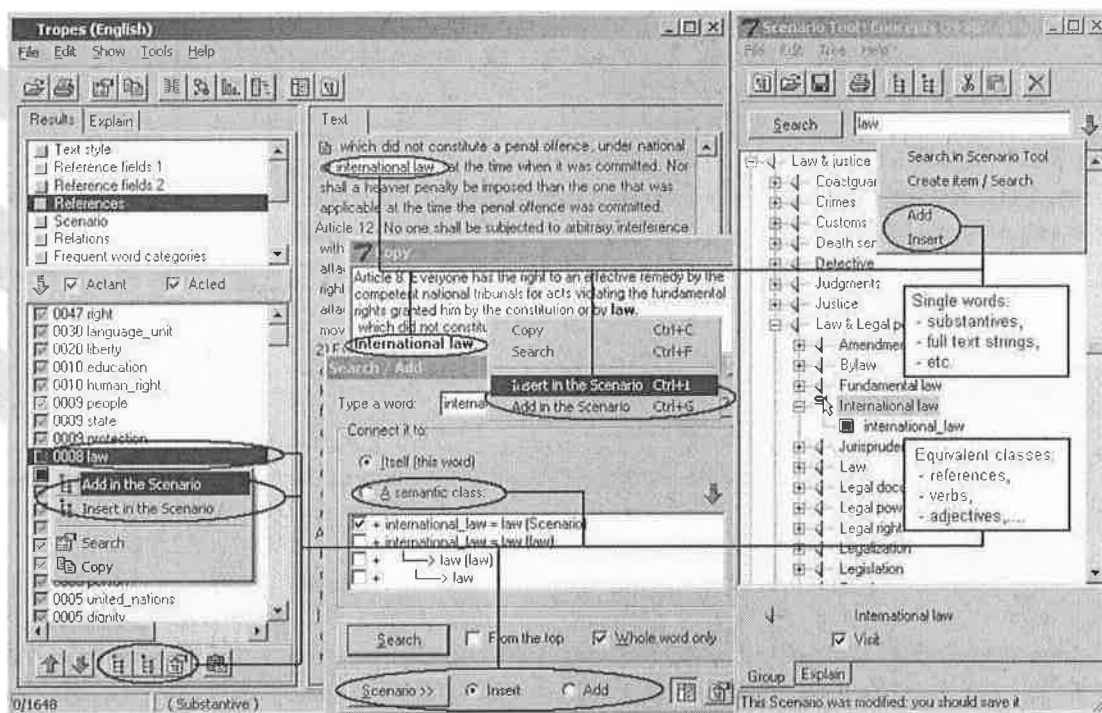


Inovație în administrație
Programul Operațional
"Dezvoltarea Capacității
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013 selected Equivalent class. The [Insert in the Scenario] command creates an entry by using the group name currently displayed in the Scenario.

Lastly, you can create a new entry in the Scenario manually from the [Search/Add] dialog box. Just type a word, select a group and press the [Scenario] button. You can choose to add a word or an Equivalent class to the Scenario. It is possible to switch from the [Insert] function to the [Add] function (and vice versa) by checking a box located at the bottom of the [Search/Add] dialog.

All the parts of the Tropes Graphic Interface contain popup menus (right-click with the mouse) which enable easy addition/insertion of words or Equivalent classes in the Scenario:



The [Save] button in the Scenario tool enable you (without quitting this dialog) to take into account the modifications to the Scenario, save them and update the display.

Most of the navigation functions in the Scenario tool are similar to those of Windows Explorer. For example, to rename a semantic group, just select it for a few seconds.¹

„The methodology of Scenario design

We suggest that you adopt the following methodology, which gives good results quickly:

1. open the Scenario tool ([Tools][Scenario] menu in the main window) and load one of the existing Scenarios ([File][Open] menu in the Scenario tool), which you then save under another name ([File][Save as] menu in the Scenario tool);

¹ Tropes Version 8.0– Reference Manual, pp. 32-33, a se vedea pe www.semantic-knowledge.com



UNIUNEA EUROPEANĂ
Fondul Social European



GUVERNUL ROMÂNIEI
Ministerul Dezvoltării Regionale
și Administrației Publice



Inovație în administrație
Programul Operațional
"Dezvoltarea Capacității
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013

2. analyze a collection of documents representative of the subject that you want to deal with;
3. display the References;
4. supplement the Scenario by adding all the words that are not classified (that is to say, are not checked in the list of References) and which are relevant to the subject of your analysis;
5. display the Relations, or the expressions proposed by the terminology extractor, and use them to classify the relevant compounds in the Scenario (for example, if you are studying the Oscar Wilde files in the example texts, you should group "Mr. Worthing" together with "Ernest" to carry out a correct analysis of the famous play);
6. you may wish to study the lists of verbs and adjectives, some of which can strengthen the Scenario;
7. apply your Scenario to the text ([File][Save] menu in the Scenario tool);
8. move on to the Scenario in the Tropes results window (using the [Show][Scenario] menu in the main window if necessary);
9. examine the result obtained by going down the list of groups in the Scenario;
10. correct misclassifications by adding them directly in the Scenario tool (for example, if "the White House" is classed among colors, you must move it into another group);
11. remove the narrative branches (which are not useful for your analysis) from your Scenario (for example, the days of the week are generally of little interest when analyzing newspaper articles);
12. interpret the result and go back to step 7, if necessary;
13. when you have checked everything, your analysis is done.

As the References are sorted in decreasing order of frequency, you can be sure that you have classified the main references in your texts when you employ this method, even if you do not incorporate all the semantic classes in the Scenario. For example, if you halt the construction of the Scenario when the References reach a frequency of two occurrences (i.e., generally not very relevant), you cannot be accused of making a hash of the analysis. You will simply have focused on the essentials, by halting the analysis at a particular moment."²

Acestea sunt principiile metodologice propuse de autorii programului. Vom mai adăuga că folosirea extractorului terminologic permite controlul riguros al terminologiei. Prin urmare, pentru construirea scenariilor necesare analizelor noastre, am folosit programul *Tropes V. 8.1* atât în limba engleză cât și în română (după caz).

PERSPECTIVA INFORMAȚIONALĂ (LINGVISTICĂ)

Așa cum menționează și autorii programului *Tropes* în „Metodologia creării unui scenariu”, la care am făcut mai sus referire, scenariile create de noi au la bază întotdeauna un corpus specific. Este evident că atunci când pretendem elaborarea unui scenariu de amploare, probabil cel mai bogat scenariu disponibil pe această temă, am avut în vedere o serie de studii preliminare terminologice.

² Idem, p. 37.



UNIUNEA EUROPEANĂ
Fondul Social European



GUVERNUL ROMÂNIEI
Ministerul Dezvoltării Regionale
și Administrației Publice



Inovație în administrație
Programul Operațional
"Dezvoltarea Capacității
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013. Crearea unor instrumente informatice de cercetare având ca temă Educația, și care să poată permite publicarea unor analize *online*, implică, pe lângă un *corpus* textual adecvat, vast și actualizat, un *thesaurus* cu navigare rapidă și fiabilă pentru tematica investigată.

Un *thesaurus* este un vocabular normalizat care implică organizarea termenilor esențiali (legați între ei prin relații semantice de tip sinonimic, asociativ și ierarhic) dintr-un anumit domeniu al cunoașterii. Acest lucru ne va permite clasificarea conținutului unui text sau unui grup de texte aparținând domeniului dat, prin evidențierea unor descriptori (cuvinte cheie) pe care specialistul îi stabilește după o analiză prealabilă a conținutului textual. Altfel spus, conținutul unui text este redus la o listă de termeni (descriptori) care să permită navigarea și recuperarea informației într-o cercetare tematică viitoare.

Un *thesaurus* electronic, de felul celui propus de noi în proiect, ambiționează însă mult mai mult.

Deși are la bază cinci *thesauri* elaborați anterior (EUROVOC, ERIC, European Education Thesaurus, TESE și Thesaurus-ul UNESCO, cu preferință pentru acesta) și o enciclopedia a educației, nu ne-am limitat doar la transcrierea terminologiei, ci am verificat și absorbit toate modalitățile prin care o idee este exprimată, confruntând cercetarea tradițională cu extragerea terminologiei din *corpus*. Altfel spus, *thesaurus* nostru electronic emerge nu doar din experiența lingvistică a echipei sau din bibliografie, ci, deopotrivă, din însăși materia-primă textuală.

Thesaurus Education (adaptat PB-EHEA) are ca intrări disponibile (a se vedea metodologia creării corpusului) o serie de termeni noi, extrași în mod automat cu ajutorul programului *Tropes*. Un astfel de demers presupune o serie de beneficii, imposibil de obținut prin indexarea manuală clasică:

- Reperarea și colectarea tuturor termenilor care exprimă aceeași idee (sinonime, echivalenți, variate libere, variante ortografice US-UK) pe baza cercetării exhaustive a documentelor. Astfel, au fost luate în considerație toate formele care apar de cel puțin două ori în două texte diferite. Menționăm că, cu cât *corpusul* este mai mare, cu atât frecvențele de sunt mai semnificative.
- Evitarea repetițiilor, a intrărilor duble. Programul nu permite ca un termen să fie, din neatenție, inclus în două serii conceptuale, așa cum am observat în mai toate materialele cercetate. Iată un exemplu de eroare, din TESE:



UNIUNEA EUROPEANĂ
Fondul Social European



GUVERNUL ROMÂNIEI
Ministerul Dezvoltării Regionale
și Administrației Publice



Inovație în administrație
Programul Operațional
"Dezvoltarea Capacității
Administrative"

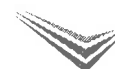
Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013

- → NT3·secondary·school·[+~]·¶¶
- → NT4·comprehensive·secondary·school¶¶
- → NT3·special·school¶¶
- → NT3·specialist·school¶¶
- → NT3·vocational·school¶¶
- → NT2·teacher·education·institution¶¶
- → NT2·training·centre¶¶
- → NT1·equipment·[+~]·¶¶
- → NT2·audiovisual·equipment¶¶
- → NT2·furniture¶¶
- → NT2·ICT·equipment·[+~]·¶¶
- → NT3·hardware·[+~]·¶¶
- → NT4·computer^o[+~]¶¶
- → NT3·server¶¶
- → NT3·software·[+~]·¶¶
- → NT4·educational·software¶¶
- → NT2·sports·equipment¶¶
- → NT1·ICT·equipment·[+~]·¶¶
- → NT2·hardware·[+~]·¶¶
- → NT3·computer·[+~]·¶¶
- → NT4·home·computer¶¶
- → NT2·server¶¶
- → NT2·software·[+~]·¶¶
- → NT3·educational·software¶¶
- → NT1·maintenance·and·repair¶¶
- → NT1·school·construction·[+~]·¶¶
- → NT2·space·arrangement·[+~]·¶¶
- → NT3·access·for·the·disabled¶¶

- Îmbogățirea *thesaurus*-ului cu termeni noi, neluați până acum în seamă, dar care apar deseori asociați termenilor introduși manual. Aceasta înseamnă, bunăoară, ca lista numelor geografice din scenariu să fie exhaustivă în raport cu textele analizate, ceea ce înseamnă un enorm câștig.
- O organizare ierarhică mai „transparentă” pentru utilizator, pe mai multe trepte de profunzime (arborescentă).
- Detașarea unor macroteme (Economie, Drept), deși acestea nu se referă în mod expres la domeniul Educației, dar care sunt păstrate pentru a putea pune în evidență relațiile dintre domenii.
- Posibilitatea organizării și reorganizării, din perspectiva utilizatorului, a tuturor serilor ierarhice, ori de câte ori va fi necesar.
- Posibilitatea degajării unui *thesaurus* permutat (reorganizat pe al doilea sau al treilea termen al unei sintagme), pentru noi cercetări.
- Posibilitatea afișării imediate a frazelor care includ termenul căutat, cu acces direct la documentul original.
- Posibilitatea îmbogățirii, revizuirii, editării și disponibilizării *online*, ca material de sprijin pentru specialiștii din domeniu.

uefiscedi

UNITATEA EXECUTIVĂ PENTRU
FINANȚAREA ÎNVĂȚĂMÂNTULUI
SUPERIOR, A CERCETĂRII,
DEZVOLTĂRII ȘI INOVĂRII



Politici publice fundamentate
în Învățământul Superior



UNIUNEA EUROPEANĂ
Fondul Social European



GUVERNUL ROMÂNIEI
Ministerul Dezvoltării Regionale
și Administrației Publice



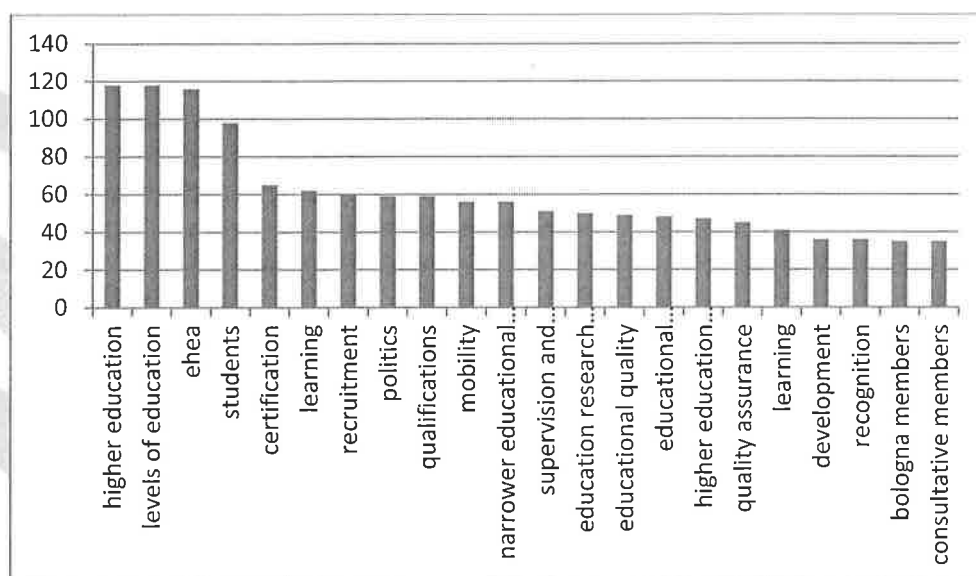
Inovație în administrație
Programul Operațional
"Dezvoltarea Capacității
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013

- Posibilitatea ca orice cercetător, independent de locul unde se află, să poată accesa pagina Internet sau să poată cere analize particulare (relații, frecvențe, statistici etc.).

Menționăm că un astfel de demers are un sporit grad de originalitate și că nu avem cunoștință de nicio altă abordare similară. În actuala sa formă, premergătoare analizelor specializate, scenariul nostru conține ceva mai mult de 5000 de termeni (nu luăm în calcul variantele).

Pentru ilustrare, vom publica graficul analizei unui mic corpus, cel al Conferințelor ministeriale ale Procesului Bologna. Pot fi vizualizați cei mai frecvenți termeni (cu peste 35 de ocurențe):



Desigur, în analizele propuse și mai ales în interpretarea rezultatelor, metodologia analitică, precum și folosirea setului de instrumente semantice, trebuie adaptate cu abilitate.



UNIUNEA EUROPEANĂ
Fondul Social European



GUVERNUL ROMÂNIEI
Ministerul Dezvoltării Regionale
și Administrației Publice



Inovație în administrație
Programul Operațional
"Dezvoltarea Capacității
Administrative"

Proiect cofinanțat din Fondul Social European, prin Programul Operațional "Dezvoltarea Capacității Administrative", în perioada 2007-2013

BIBLIOGRAFIE

Encyclopedia of Education, 2nd edition, Macmillan Reference, 2004

ERIC Thesaurus, disponibil pe <http://www.csa.com/factsheets/supplements/ericthes.php>

European Education Thesaurus, disponibil pe <http://www.freethesaurus.info/redined/en/index.php>

Eurovoc, the EU's multilingual thesaurus, disponibil pe www.eurovoc.europa.eu/

TESE – Thesaurus Europäischer Bildungssysteme, disponibil pe
http://eacea.ec.europa.eu/education/eurydice/documents/tese/pdf/tesede_005_alphabetic.pdf

Tropes Version 8.0 – Reference Manual, disponibil pe www.semantic-knowledge.com

UNESCO IBE: Education Thesaurus, 6th edition, 2007, disponibil pe
<http://www.ibe.unesco.org/en/services/online-materials/publications/ibe-education-thesaurus.html>