Academic Performance Prediction Using Chance Discovery from Online Discussion Forums

COMPSAC 2016 -

Preliminary Results Presentation

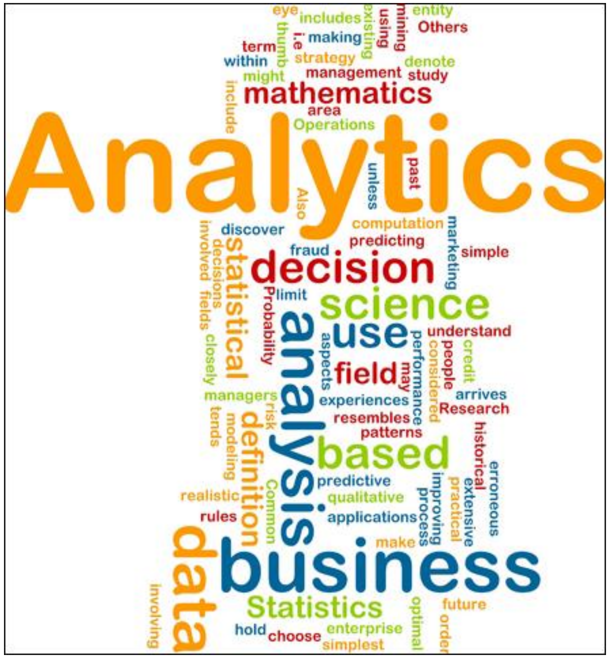
By Dr. Gary. K. W. Wong and Simon Y. K. Li

Outline

- Introduction and Motivation
- Project Objectives
- Prior Works
- Experiment Design
- Preliminary Results
- Conclusion and Discussion
- Potential Contribution
- Follow-up

Introduction and Motivation

- E-learning systems have become more popular recently in the higher education sector producing a **large volume of data** created by students and teachers inside via collaborative online discussion forums, which are **difficult to analyze without using data mining technologies**.



- The forums contain a lot of useful data for **educational data mining** to extract **insightful patterns** so that educators can better understand the thinking patterns of students during the learning process.

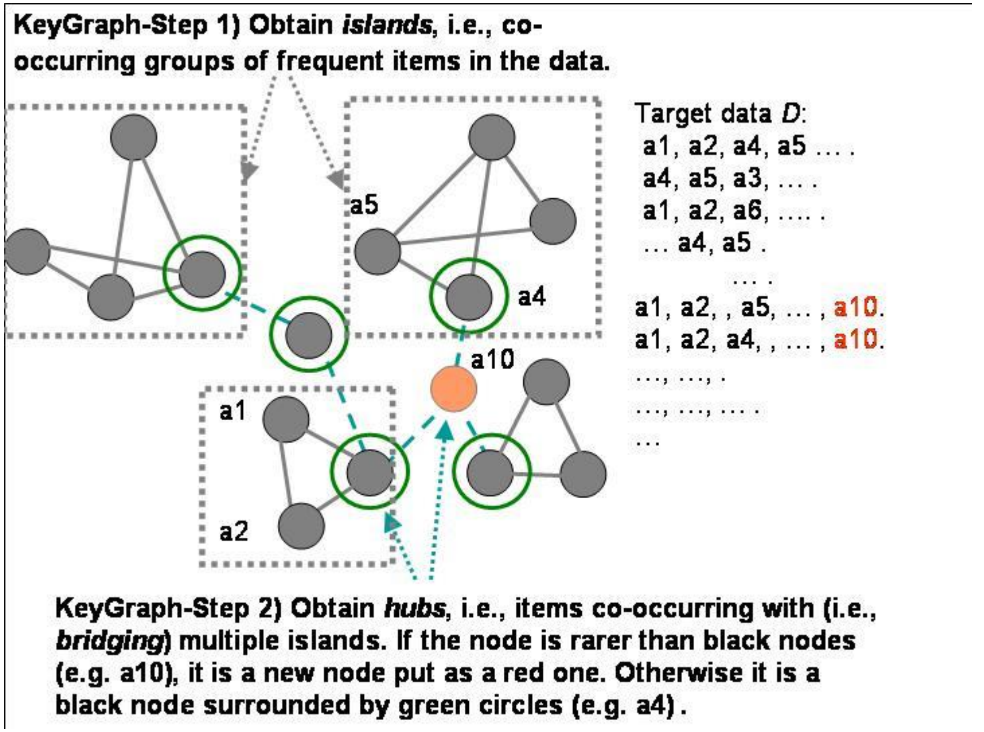
Project Objectives

- Identify / develop innovative **analytical methods and tools** facilitating teachers' assessment to evaluate the students learning in the online learning environment to determine if students are able to meet **learning objective** and / or **generate new knowledge** beyond the expected learning activities (serendipitous / accidental learning).
- Present **our preliminary research** findings on educational data mining using student discussion forums for future research planning.
- **Adaptive pedagogical design** can be made for teaching and learning.

香港教育大學
The Education University of Hong Kong

Prior Works - KeyGraph

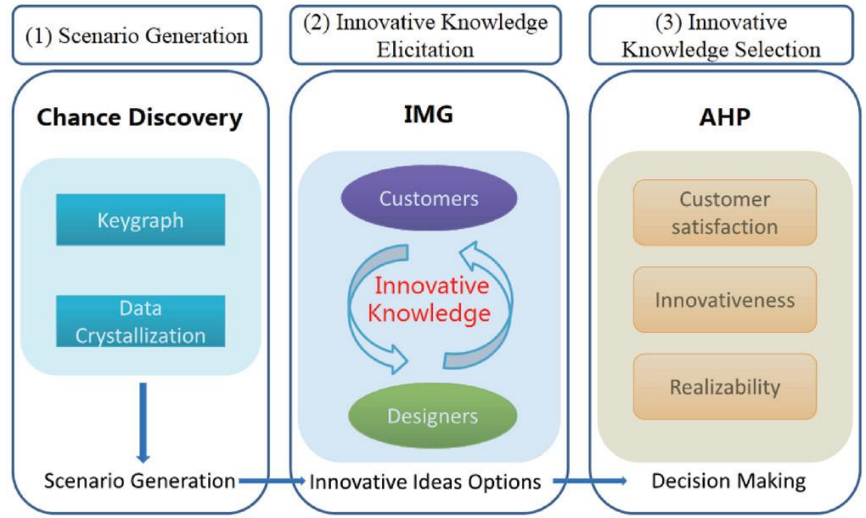
- The black nodes and black links represent the **frequent items and their co-occurrences**, implying an *established trend in the data*.
- Black nodes and links form **clusters** representing ***concepts***.
- The red nodes and red links *connect multiple clusters* or some phenomena such as transition of events from one to other clusters.
- Red nodes can be regarded as *chances*.
- Keygraphs are generated to visualize "**scenarios**" to understand the situation.



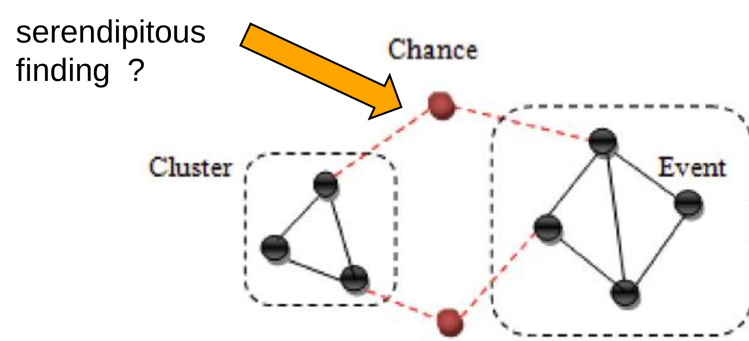
Source: Ohsawa, Y., Benson, N. E., & Yachida, M. (1998, April). KeyGraph: Automatic indexing by co-occurrence graph based on building construction metaphor.

Prior Works - Chance Discovery (CD)

- Three stages: Scenario Generation, IMG, and AHP.
- KeyGraph is the algorithm used to **generate scenarios** for **finding chances** from document to (a) ***valuable result***(s).
- IMG aims at evaluating the chances using human intervention (with expert / prepared minds) to identify values, which is enhanced by using the third stage AHP.
- Reference: Wang, H., Ohsawa, Y., & Nishihara, Y. (2013). A system method to elicit innovative knowledge based on chance discovery for innovative product design.

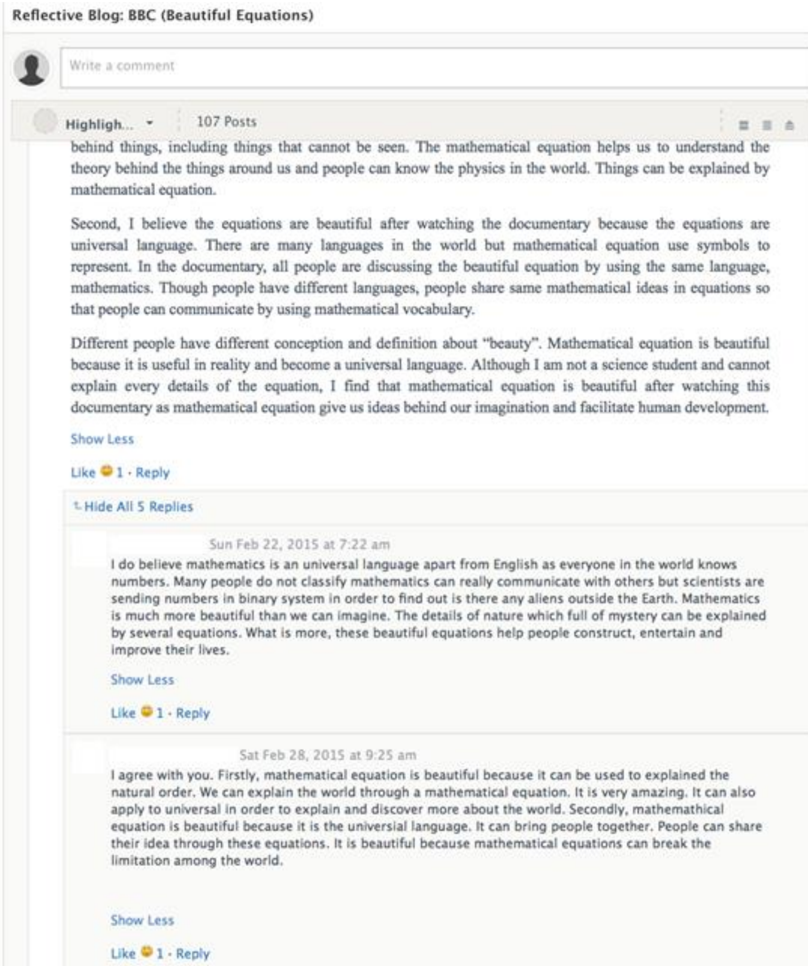


Scenario diagram generated by KeyGraph



Experiment Design

- In this project, a total 24 undergraduate students in the HKIEd from the General Education course called "***Technology, Entertainment and Mathematics***" have been sampled for this preliminary experiment.
- One of the course requirements was to complete a *reflective posting* on an online discussion forum in Schoology.
- They were asked to watch a BBC documentary film called "***Beautiful Equations***" and other ***selected movies***.
- Afterwards they posted their reflections in the forum. Each student was also required to comment on three self-selected peers, which were extracted in our experiment as text files for analysis.
- There was a total of 110 posts.
- We copied and pasted the posts to text files for storage and further analysis.


Reflective Blog: BBC (Beautiful Equations)

Write a comment

Highligh... 107 Posts

behind things, including things that cannot be seen. The mathematical equation helps us to understand the theory behind the things around us and people can know the physics in the world. Things can be explained by mathematical equation.

Second, I believe the equations are beautiful after watching the documentary because the equations are universal language. There are many languages in the world but mathematical equation use symbols to represent. In the documentary, all people are discussing the beautiful equation by using the same language, mathematics. Though people have different languages, people share same mathematical ideas in equations so that people can communicate by using mathematical vocabulary.

Different people have different conception and definition about "beauty". Mathematical equation is beautiful because it is useful in reality and become a universal language. Although I am not a science student and cannot explain every details of the equation, I find that mathematical equation is beautiful after watching this documentary as mathematical equation give us ideas behind our imagination and facilitate human development.

Show Less

Like 1 · Reply

Hide All 5 Replies

Sun Feb 22, 2015 at 7:22 am

I do believe mathematics is an universal language apart from English as everyone in the world knows numbers. Many people do not classify mathematics can really communicate with others but scientists are sending numbers in binary system in order to find out is there any aliens outside the Earth. Mathematics is much more beautiful than we can imagine. The details of nature which full of mystery can be explained by several equations. What is more, these beautiful equations help people construct, entertain and improve their lives.

Show Less

Like 1 · Reply

Sat Feb 28, 2015 at 9:25 am

I agree with you. Firstly, mathematical equation is beautiful because it can be used to explained the natural order. We can explain the world through a mathematical equation. It is very amazing. It can also apply to universal in order to explain and discover more about the world. Secondly, mathemathical equation is beautiful because it is the universial language. It can bring people together. People can share their idea through these equations. It is beautiful because mathematical equations can break the limitation among the world.

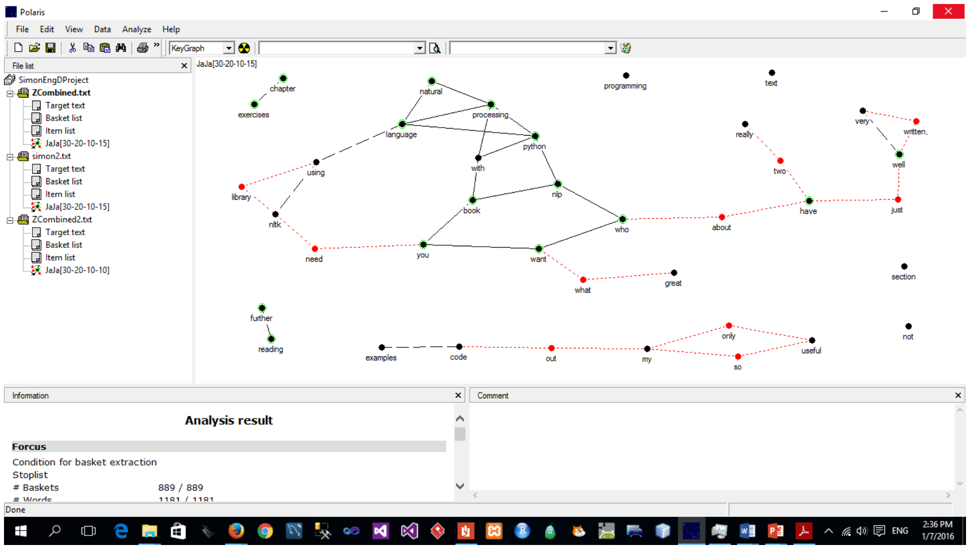
Show Less

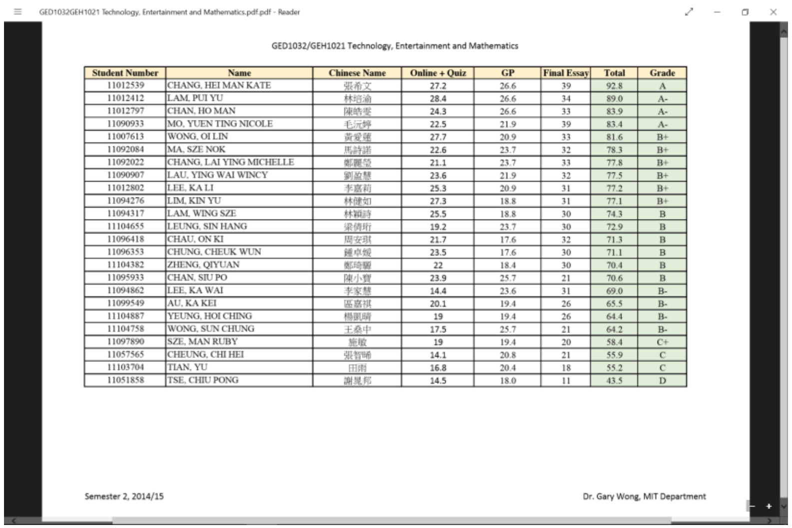
Like 1 · Reply

香港教育大學
The Education University of Hong Kong

Experiment Design

- A software tools called “Polaris” from Ohsawa Laboratory was used for mining text from the following sources.
- Sources of Data (text format):
  - Online reflective discussion forum, etc.
  - Students’ final grade
- Group the discussion posts by <u>their final grades</u> of the module, e.g. A, B, C, D, etc.
- Performed analysis using KeyGraph to generate the visual patterns to identify:
  - The formulation of *<u>key concepts</u>* from black nodes and links
  - *<u>chances</u>* (red nodes and links) for the purposes of decision making and planning in the associated areas above.
  - *<u>See whether there are patterns of back / red nodes / links across different grades for future predictive applications</u>*





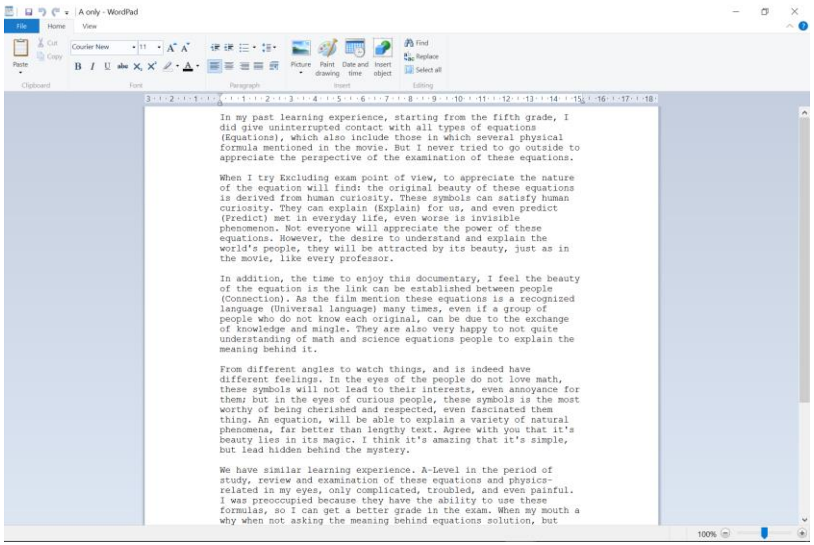
GED1032/GEH1021 Technology, Entertainment and Mathematics

Student Number	Name	Chinese Name	Online + Quiz	GP	Final Essay	Total	Grade
11012539	CHANG, HEI MAN KATE	張希文	27.2	26.6	39	92.8	A
11012412	LAM, PUI YU	林培渝	28.4	26.6	34	89.0	A-
11012797	CHAN, HO MAN	陳皓雯	24.3	26.6	33	83.9	A-
11090933	MO, YUEN TING NICOLE	毛沅婷	22.5	21.9	39	83.4	A-
11007613	WONG, OI LIN	黃愛蓮	27.7	20.9	33	81.6	B+
11092084	MA, SZE NOK	馬詩諾	22.6	23.7	32	78.3	B+
11092022	CHANG, LAI YING MICHELLE	鄭麗瑩	21.1	23.7	33	77.8	B+
11090907	LAU, YING WAI WINCY	劉盈慧	23.6	21.9	32	77.5	B+
11012802	LEE, KA LI	李嘉莉	25.3	20.9	31	77.2	B+
11094276	LIM, KIN YU	林健如	27.3	18.8	31	77.1	B+
11094317	LAM, WING SZE	林穎詩	25.5	18.8	30	74.3	B
11104655	LEUNG, SIN HANG	梁倩珩	19.2	23.7	30	72.9	B
11096418	CHAU, ON KI	周安琪	21.7	17.6	32	71.3	B
11096353	CHUNG, CHEUK WUN	鍾卓媛	23.5	17.6	30	71.1	B
11104382	ZHENG, QIYUAN	鄭琦[illegible]	22	18.4	30	70.4	B
11095933	CHAN, SIU PO	陳小寶	23.9	25.7	21	70.6	B
11094862	LEE, KA WAI	李家慧	14.4	23.6	31	69.0	B-
11099549	AU, KA KEI	區嘉祺	20.1	19.4	26	65.5	B-
11104887	YEUNG, HOI CHING	楊凱晴	19	19.4	26	64.4	B-
11104758	WONG, SUN CHUNG	王燊中	17.5	25.7	21	64.2	B-
11097890	SZE, MAN RUBY	施敏	19	19.4	20	58.4	C+
11057565	CHEUNG, CHI HEI	張智晞	14.1	20.8	21	55.9	C
11103704	TIAN, YU	田雨	16.8	20.4	18	55.2	C
11051858	TSE, CHIU PONG	謝晁邦	14.5	18.0	11	43.5	D

Semester 2, 2014/15 Dr. Gary Wong, MIT Department

香港教育大學
The Education University of Hong Kong

Input Text Files



A only - WordPad

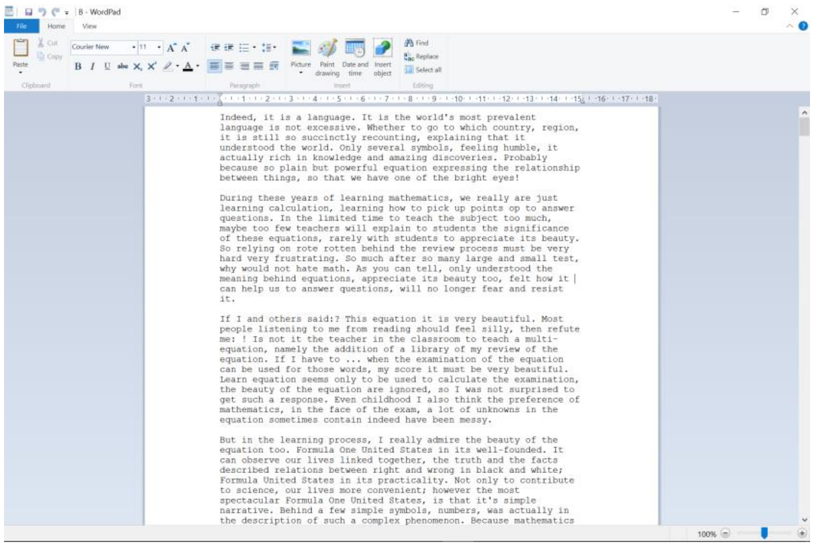
In my past learning experience, starting from the fifth grade, I did give uninterrupted contact with all types of equations (Equations), which also include those in which several physical formula mentioned in the movie. But I never tried to go outside to appreciate the perspective of the examination of these equations.

When I try Excluding exam point of view, to appreciate the nature of the equation will find: the original beauty of these equations is derived from human curiosity. These symbols can satisfy human curiosity. They can explain (Explain) for us, and even predict (Predict) met in everyday life, even worse is invisible phenomenon. Not everyone will appreciate the power of these equations. However, the desire to understand and explain the world's people, they will be attracted by its beauty, just as in the movie, like every professor.

In addition, the time to enjoy this documentary, I feel the beauty of the equation is the link can be established between people (Connection). As the film mention these equations is a recognized language (Universal language) many times, even if a group of people who do not know each original, can be due to the exchange of knowledge and mingle. They are also very happy to not quite understanding of math and science equations people to explain the meaning behind it.

From different angles to watch things, and is indeed have different feelings. In the eyes of the people do not love math, these symbols will not lead to their interests, even annoyance for them; but in the eyes of curious people, these symbols is the most worthy of being cherished and respected, even fascinated them thing. An equation, will be able to explain a variety of natural phenomena, far better than lengthy text. Agree with you that it's beauty lies in its magic. I think it's amazing that it's simple, but lead hidden behind the mystery.

We have similar learning experience. A-Level in the period of study, review and examination of these equations and physics-related in my eyes, only complicated, troubled, and even painful. I was preoccupied because they have the ability to use these formulas, so I can get a better grade in the exam. When my mouth a why when not asking the meaning behind equations solution, but



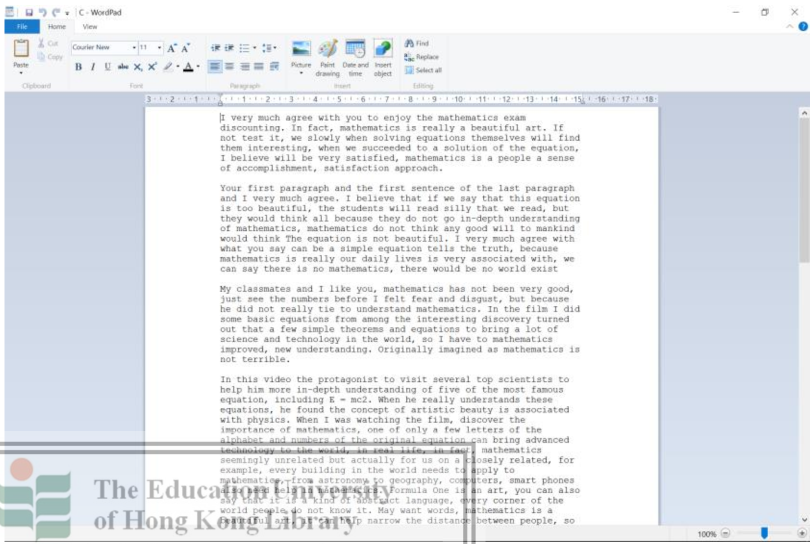
B - WordPad

Indeed, it is a language. It is the world's most prevalent language is not excessive. Whether to go to which country, region, it is still so succinctly recounting, explaining that it understood the world. Only several symbols, feeling humble, it actually rich in knowledge and amazing discoveries. Probably because so plain but powerful equation expressing the relationship between things, so that we have one of the bright eyes!

During these years of learning mathematics, we really are just learning calculation, learning how to pick up points op to answer questions. In the limited time to teach the subject too much, maybe too few teachers will explain to students the significance of these equations, rarely with students to appreciate its beauty. So relying on rote rotten behind the review process must be very hard very frustrating. So much after so many large and small test, why would not hate math. As you can tell, only understood the meaning behind equations, appreciate its beauty too, felt how it can help us to answer questions, will no longer fear and resist it.

If I and others said:? This equation it is very beautiful. Most people listening to me from reading should feel silly, then refute me: ! Is not it the teacher in the classroom to teach a multi-equation, namely the addition of a library of my review of the equation. If I have to ... when the examination of the equation can be used for those words, my score it must be very beautiful. Learn equation seems only to be used to calculate the examination, the beauty of the equation are ignored, so I was not surprised to get such a response. Even childhood I also think the preference of mathematics, in the face of the exam, a lot of unknowns in the equation sometimes contain indeed have been messy.

But in the learning process, I really admire the beauty of the equation too. Formula One United States in its well-founded. It can observe our lives linked together, the truth and the facts described relations between right and wrong in black and white; Formula United States in its practicality. Not only to contribute to science, our lives more convenient; however the most spectacular Formula One United States, is that it's simple narrative. Behind a few simple symbols, numbers, was actually in the description of such a complex phenomenon. Because mathematics



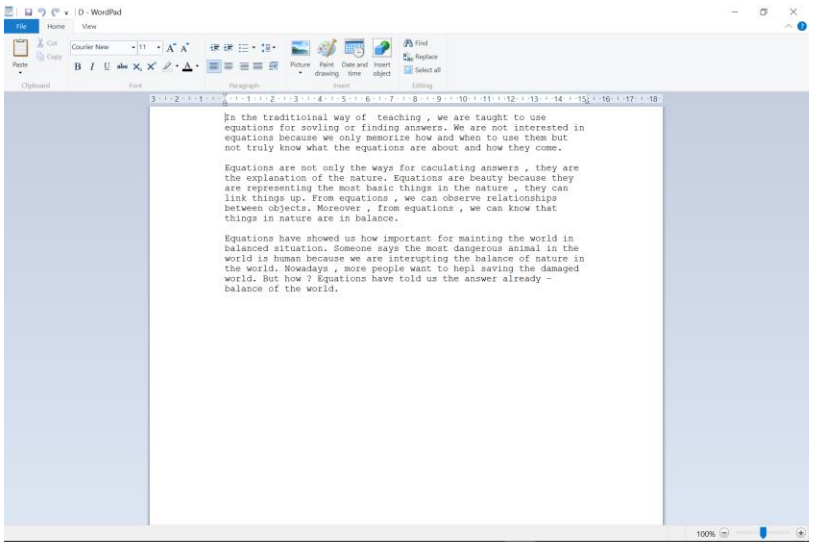
C - WordPad

I very much agree with you to enjoy the mathematics exam discounting. In fact, mathematics is really a beautiful art. If not test it, we slowly when solving equations themselves will find them interesting, when we succeeded to a solution of the equation, I believe will be very satisfied, mathematics is a people a sense of accomplishment, satisfaction approach.

Your first paragraph and the first sentence of the last paragraph and I very much agree. I believe that if we say that this equation is too beautiful, the students will read silly that we read, but they would think all because they do not go in-depth understanding of mathematics, mathematics do not think any good will to mankind would think The equation is not beautiful. I very much agree with what you say can be a simple equation tells the truth, because mathematics is really our daily lives is very associated with, we can say there is no mathematics, there would be no world exist

My classmates and I like you, mathematics has not been very good, just see the numbers before I felt fear and disgust, but because he did not really tie to understand mathematics. In the film I did some basic equations from among the interesting discovery turned out that a few simple theorems and equations to bring a lot of science and technology in the world, so I have to mathematics improved, new understanding. Originally imagined as mathematics is not terrible.

In this video the protagonist to visit several top scientists to help him more in-depth understanding of five of the most famous equation, including E = mc2. When he really understands these equations, he found the concept of artistic beauty is associated with physics. When I was watching the film, discover the importance of mathematics, one of only a few letters of the alphabet and numbers of the original equation can bring advanced technology to the world, in real life, in fact, mathematics seemingly unrelated but actually for us on a closely related, for example, every building in the world needs to apply to mathematics, from astronomy to geography, computers, smart phones also need help in mathematics. Formula One is an art, you can also say that it is a kind of abstract language, every corner of the world people do not know it. May want words, mathematics is a beautiful art, it can help narrow the distance between people, so



D - WordPad

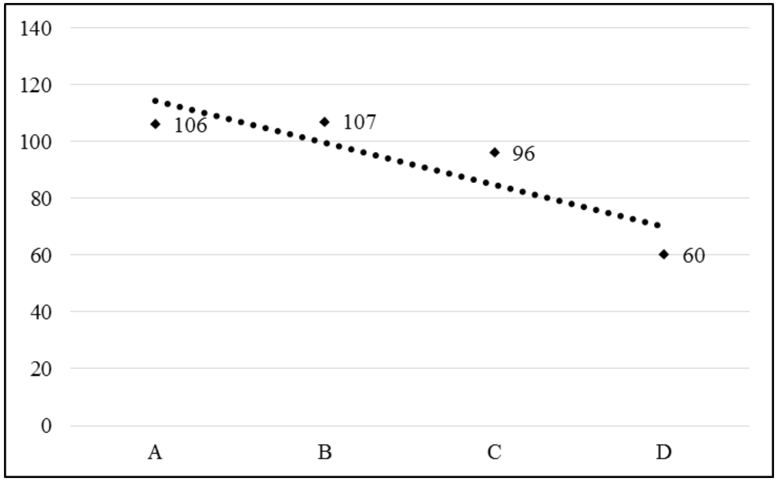
In the traditioinal way of teaching , we are taught to use equations for sovling or finding answers. We are not interested in equations because we only memorize how and when to use them but not truly know what the equations are about and how they come.

Equations are not only the ways for caculating answers , they are the explanation of the nature. Equations are beauty because they are representing the most basic things in the nature , they can link things up. From equations , we can observe relationships between objects. Moreover , from equations , we can know that things in nature are in balance.

Equations have showed us how important for mainting the world in balanced situation. Someone says the most dangerous animal in the world is human because we are interupting the balance of nature in the world. Nowadays , more people want to hepl saving the damaged world. But how ? Equations have told us the answer already - balance of the world.

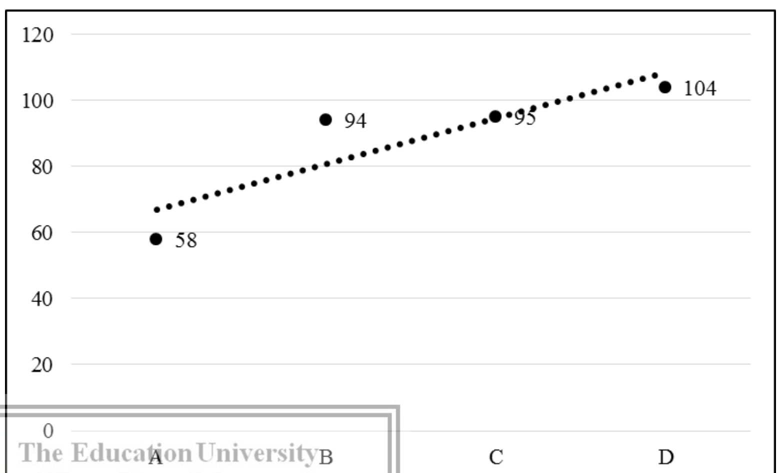
Preliminary results (Black nodes / links - concepts)

No. of Black Nodes vs. Academic Grades



A declining trend of having less black nodes was observed

No. of Black Links vs. Academic Grades

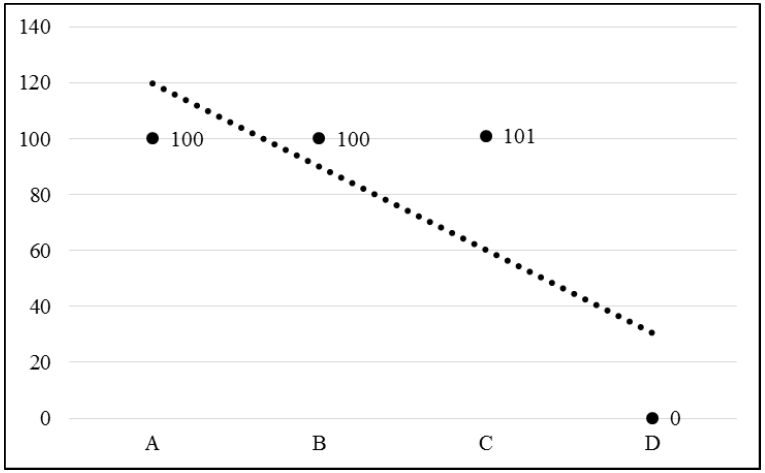


A rising trend of having more black links.

香港教育大學
The Education University of Hong Kong

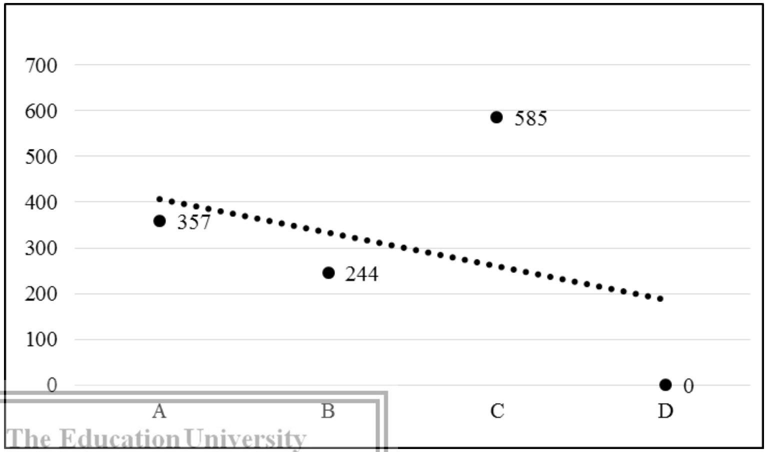
Preliminary results (Red nodes / links - chance)

No. of Red Nodes vs. Academic Grades



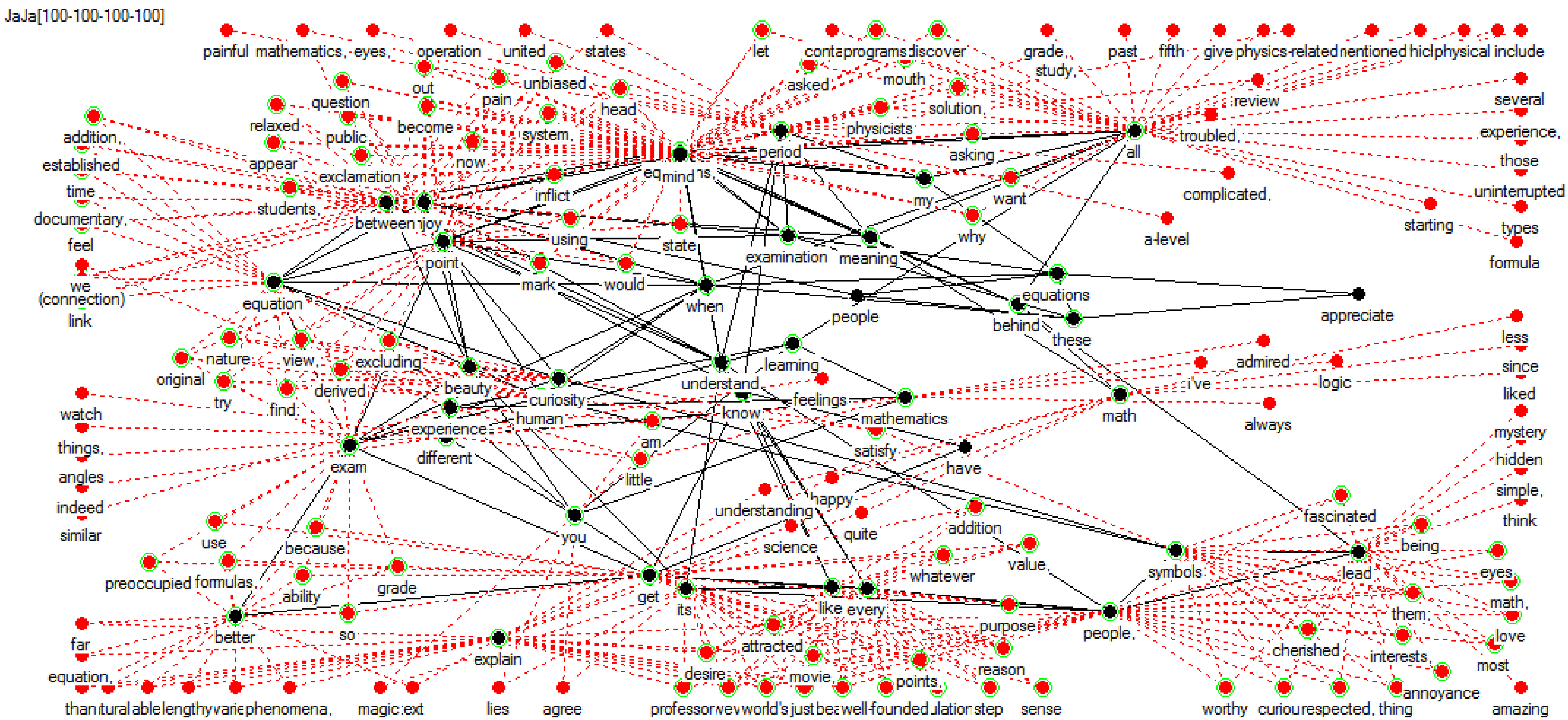
The number of red nodes for the grade D dropped to zero while for other grade the differences were minimal.

No. of Red Links vs. Academic Grades



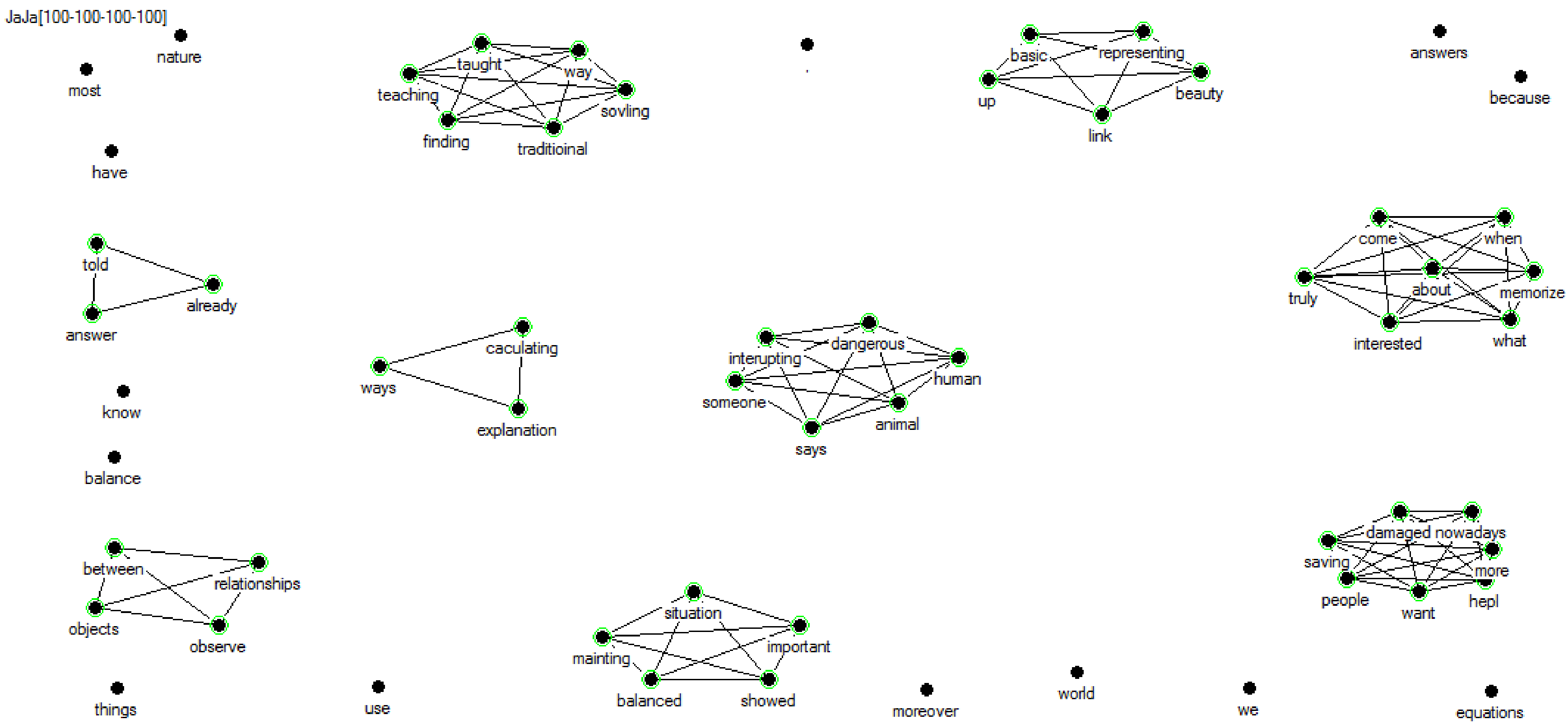
The number of red links was generally on the declining trend across the grades although the number grew sharply for grade C

Preliminary Results – Student with A grade



香港教育大學
The Education University of Hong Kong

Preliminary Results – student with D grade



Conclusion and Discussion

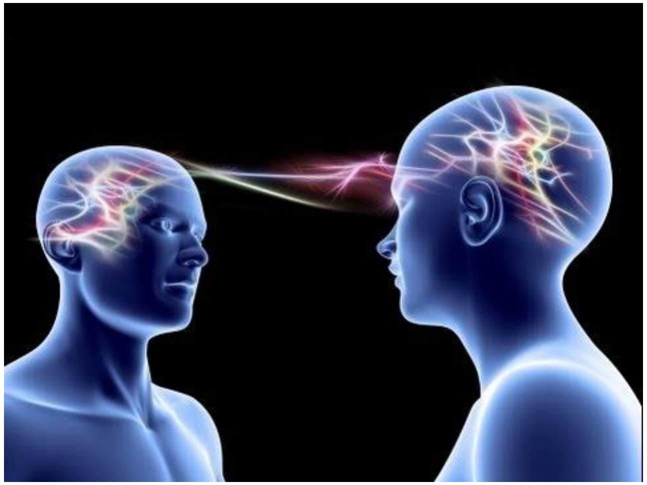
- The above results can **potentially provide insights** into the **performance prediction** of students.
- The students who can score **better grades** (e.g. "A", "A-", "B", etc.) would usually have the tendency to contribute **more in-depth contents** in their posts and hence **creating more "chances" and linkages** to other concepts from their posts, as identified as the red nodes and links in the above KeyGraphs shown.
- While those of the students who scored bad grades (e.g. "D") contributed **isolated facts with little / no smooth connection or transition from a concept to another.**
- The **numbers of "chances" and linkages** would have the potential to be used as an indicator to **predict future student performance** in some **general education modules**,.
- Having a lot of **black nodes and links** alone may not ensure a good and presentable organization of contents, which is also an important factor to deliver a good assay.
- Similarly, a lot of red nodes and links may not really mean that a lot of chances have been explored. Instead, this may just indicate to have noise in the posts. **Poor written presentation may be deduced.** The writing style may be an important area for the markers to determine grades of a written assay.
- if the student is going to study some other **non-liberal modules**, the situation may look completely different since some **more concrete** (*rather than liberal or lateral*) viewpoints would be needed. *The above patterns explored may have totally different interpretations.*
- *Therefore, merely depending on the numbers of the black / red nodes and links may not be sufficient enough without considering the context in which the contents are produced.*
- However, once those patterns can be observed, they can be accumulated and further studied to become **generalized rules**.
- **Machine learning algorithms** (e.g. classification, association, clustering, etc.)] can then be applied in our project later to explore the combinations of those patterns to formulate rules.

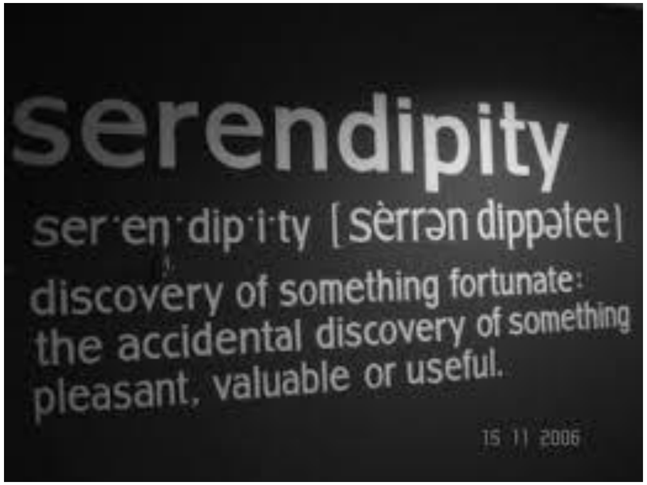
Potential Contribution

- Teachers can **better understand the patterns of thinking of students** during the learning process.

- The **assessment of students** can be facilitated through a **systematic approach** with effective **pedagogic changes** for particular students through the learning process to optimize their learning outcome

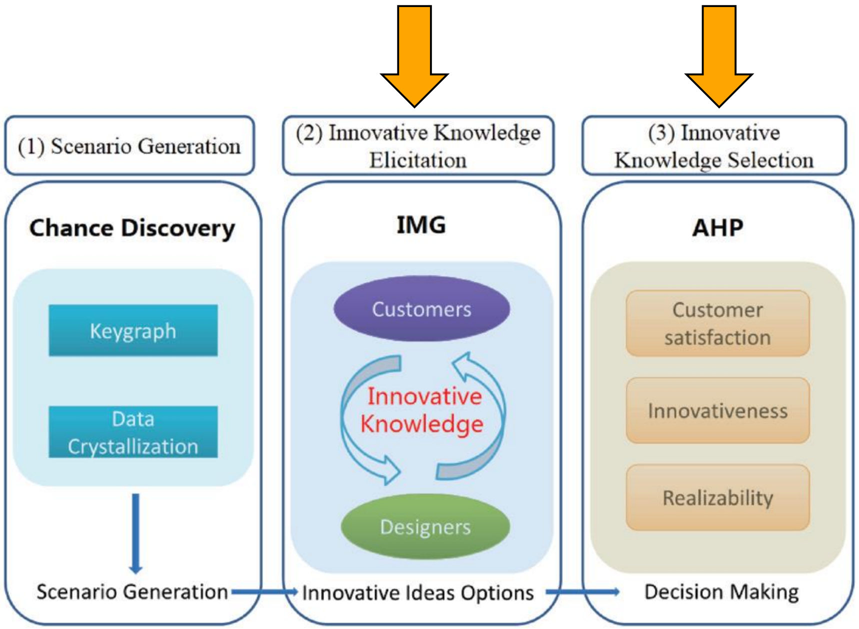
- There seems a huge potential of unexpected learning discoveries – serendipity, in collaborative online learning.





Follow-up

- Explore different perspectives to improve the analysis like
  - gender,
  - age,
  - sample sizes,
  - multiple academic subjects,
  - writing styles / language competence,
  - and many other qualitative factors.
- Revise the stop words list to filter noise.
- Further cleanse the sources of text to fix some conversion errors.
- Introduce semantics technologies into matching (dis)similarity
- Perform the KeyGraph analysis again or select more powerful text mining tools.
- Proceed to the IMG and AHP stages of CD for evaluation of the chances already spotted.



References

- André, P., Teevan, J., & Dumais, S. T. (2009, October). Discovery is never by chance: designing for (un) serendipity. In Proceedings of the seventh ACM conference on Creativity and cognition (pp. 305-314). ACM.
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. Assessment & Evaluation in Higher Education, 31(4), 399-413.
- Buchem, I. (2011). Serendipitous learning: Recognizing and fostering the potential of microblogging. Form@ re-Open Journal per la formazione in rete, 11(74), 7-16.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. Review of educational research, 65(3), 245-281.
- Delgado Kloos, C., Ibanez-Espiga, M. B., Fernández-Panadero, C., Munoz-Merino, P. J., Estevez-Ayres, I., Crespo-Garcia, R. M., ... & Perez-Sanagustin, M. (2014, October). A multidimensional analysis of trends in educational technology. In Frontiers in Education Conference (FIE), 2014 IEEE (pp. 1-4). IEEE.
- Eynon, R. (2013). The rise of Big Data: what does it mean for education, technology, and media research?. Learning, Media and Technology, 38(3), 237-240.
- Ferguson, R. (2012). Learning analytics: drivers, developments and challenges. International Journal of Technology Enhanced Learning, 4(5), 304-317.
- Foster, A., & Ford, N. (2003). Serendipity and information seeking: an empirical study. Journal of Documentation, 59(3), 321-340.
- Gaeta, M., Loia, V., Mangione, G. R., Miranda, S., & Orciuoli, F. (2014). Unlocking serendipitous learning by means of social Semantic web. In CSEDU 2014-In proceedings of the 6th international conference on computer supported education (Vol. 1, pp. 285-292).
- Gundecha, P. and Liu, H. (2012). Mining social media: A brief introduction. P.Mirchandani (Ed.) INFORMS TutORials in Operations Research, Vol. 9. Hanover:MD, INFORMS.
- Han, J., Kamber, M., & Pei, J. (2011). Data mining: concepts and techniques: concepts and techniques. Elsevier.
- Kop, R. (2012). The unexpected connection: Serendipity and human mediation in networked learning. Journal of Educational Technology & Society, 15(2), 2-11.

References

- Lehmann, T., Hähnlein, I., & Ifenthaler, D. (2014). Cognitive, metacognitive and motivational perspectives on preflection in self-regulated online learning. Computers in Human Behavior, 32, 313-323.
- Maeno, Y., & Ohsawa, Y. (2007). Human–computer interactive annealing for discovering invisible dark events. Industrial Electronics, IEEE Transactions on, 54(2), 1184-1192.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). Introduction to information retrieval (Vol. 1, p. 496). Cambridge: Cambridge university press.
- Mazza, R., & Milani, C. (2004, November). Gismo: a graphical interactive student monitoring tool for course management systems. In TEL’04 Technology Enhanced Learning’04 International Conference (pp. 18-19).
- Merton, R. K., & Barber, E. (2006). The travels and adventures of serendipity: A study in sociological semantics and the sociology of science. Princeton University Press.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. Studies in higher education, 31(2), 199-218.
- Ohsawa, Y. (2005). Data crystallization: chance discovery extended for dealing with unobservable events. New mathematics and natural computation, 1(03), 373-392.
- Ohsawa, Y., Benson, N. E., & Yachida, M. (1998, April). KeyGraph: Automatic indexing by co-occurrence graph based on building construction metaphor. In Research and Technology Advances in Digital Libraries, 1998. ADL 98. Proceedings. IEEE International Forum on (pp. 12-18). IEEE.
- Ohsawa, Y., & Fukuda, H. (2002). Chance discovery by stimulated groups of people. Application to understanding consumption of rare food. Journal of contingencies and crisis management, 10(3), 129-138.
- Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. Computers & Education, 51(1), 368-384.
- Slade, S., & Galpin, F. (2012, April). Learning analytics and higher education: ethical perspectives. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 16-17). ACM.
- Shum, S. B., & Ferguson, R. (2012). Social learning analytics. Journal of educational technology & society, 15(3), 3-26.

References

- Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning Analytics in a data-rich context. Computers in Human Behavior, 47, 157-167.
- Toms, E. G. (2000, December). Serendipitous Information Retrieval. In DELOS Workshop: Information Seeking, Searching and Querying in Digital Libraries.
- Wang, H., & Ohsawa, Y. (2011). iChance: a web-based innovation support system for business intelligence. International Journal of Organizational and Collective Intelligence (IJOCI), 2(4), 48-61.
- Wang, H., Ohsawa, Y., & Nishihara, Y. (2013). A system method to elicit innovative knowledge based on chance discovery for innovative product design. Multidisciplinary Studies in Knowledge and Systems Science, 153.
- West, D. M. (2012). Big data for education: Data mining, data analytics, and web dashboards. Governance Studies at Brookings, 1-10.
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. Higher education, 45(4), 477-501.
- Zhang, D., Liu, Y., & Si, L. (2011, August). Serendipitous learning: learning beyond the predefined label space. In Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1343-1351). ACM.

End of Presentation – Thank You