A method to convert learning history into data for learning path study in sub-Saharan Africa

Sy Khalifa, Naka Gotoda, Rihito Yaegashi & Toshihiro Hayashi
Kagawa University
s19d621@stu.kagawa-u.ac.jp

Ⅰ. Introduction

The global focus on the United Nations' Sustainable Development Goals (SDGs) has prompted African countries to try to improve their societies. Special attention has been paid to the topic of education related to SDG 4. As of 2019, many countries in sub-Sahara Africa (hereafter abbreviated to SSA) were unlikely to meet their goals, despite their best efforts.

E-Learning has great potential to help SSA states succeed in reaching their goals of SDG4, particularly if learning paths or pathway generation is adapted to the learners and their environment. Jih stated, "Learning pathways also reveal learning traits while learners traverse any interactive environment" (Jih, 1996). In other words, while learning is a lifelong journey, a learning path is the specific route a person takes to reach a desired outcome, knowledge, or achievement.

A properly constructed learning path can provide effective knowledge and skill acquisition and reduce the skill mismatch. The European Union Commission defines skill mismatch as, "a situation of imbalance in which the level or type of skills available does not correspond to labor market needs" (EU Commission, 2020). Skill mismatch is also a serious problem for young professionals including recent graduates in SSA, particularly those in the Information and Communication Technologies (ICT) sector. With considering skill mismatch, the learning path must be customized for each young professional based on her/his learning history. Usually, a lot of data as learning history are needed to generate a proper learning path. However, such a big data approach for data collection is not suitable for SSA due to cultural, educational, and even linguistic diversities and computer resources. We propose an alternative data collection method to use a limited set of learning history which consists of First-Hand Knowledge, Official Transcript and Record, and Log Data in this paper.

Ⅱ. Educational situation and the use of e-Learning in African countries

With Africa containing 44 of the 79 most underdeveloped nations in the world, it also holds 57% of the world's poor. SDG 4, which seeks to "ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" (The Sustainable Development Goals Center for Africa, 2019), has been particularly difficult for many SSA countries to achieve.

Although "education for all" might be obtainable for many developed nations, SSA countries have found it to be consistently out of reach. The outbreak of COVID-19 and its subsequent extended school closures only proved to perpetuate turbulent preexisting conditions such as frequent teacher strikes, civil unrest, and long-term power outages (Balogun & Ahlan, 2015; Kituyi & Tusubira, 2013; Adelakun & Omolola, 2020). Trained teachers in primary education were estimated

to be at 67.9% in 2019. The effect of these chaotic conditions was evident in September 2021, when the UNESCO Institute of Statistics placed literacy in SSA countries at under 66%. To compensate for such unstable learning conditions, many students seeking to attend university have turned to e-Learning for technical skill acquisition instead of traditionally formatted education.

Some e-Learning services can provide their users with customized learning paths, but the system requires a lot of data to generate them. Collecting the data is often far easier for Western users than for those in SSA. In addition, such data and their subsequently obtained learning paths generally do not have much relevance with current educational and economical situations in SSA countries. Thus, SSA students need learning paths that fit their unique situation and environment (Balogun & Ahlan, 2015; Kituyi & Tusubira, 2013). If their projected learning paths were made more accessible and relevant, it is possible that the impact of learning interruptions could decrease.

It is worth stating that while some might argue that e-Learning is not practical in an area that struggles with infrastructure challenges, it is encouraging to note the progress of the mobile economy in SSA. Mobile device internet penetration reached 46% of the region's population by the end of 2020, a 4% increase (or 20 million users) from the year before (GSMA, 2021). Therefore, SSA students would indeed be able to access their customized learning path platforms.

Ⅲ. Data collection for learning path generation

The purpose of our research is to find a method to collect proper learning history to create learning paths for young professionals in SSA. A lot of information about learning history can be obtained from log data in e-Learning platforms such as Moodle (Moodle documentation, 2020).

The typical data collection method for generating learning paths often done in developed nations is based on the big data approach with the concept that learning history equals log data. However, this method is not appropriate for SSA, because the log data are often scarce and hard to access. Therefore, other information must be used to formulate proper individual learning history. The target users of this research are young professionals in SSA, so we can focus on the period from their graduation at an educational institution to becoming young professionals. We think individual learning history can be formulated by collecting information about learning related to the period.

Ⅳ. Data collection method

To best understand how learning paths relate to real-world skills, we limited our participant selection to recent graduates and young professionals working in the ICT sector. Seeking as comprehensive an understanding as possible regarding one's current learning status, we selected three main sources of data:

1. First-Hand Knowledge
2. Official Transcripts and Records
3. Log Data

These data sources, while diverse, are processed to produce usable training data for learning path generation.

1 - First-Hand Knowledge; Written questionnaires and verbal interviews are used to inquire about

the subject's experience and skills. The questions are intentionally short and left open-ended to allow the respondent freedom in answering. The questions asked are:

- Tell us your occupation name and characteristics.
- What skills did you have when getting your job?
- What skills did you have to acquire during in-house training?

Their current occupation is asked to see if there is a direct correlation between what they studied in school and their vocation. Often interviewees would describe the same course using different names. A learning data collection tool must be able to understand the association between relatively interchangeable subjects' terms. Revealing correlations between these subjects with different names but similar contents helps to expand the default learning path with skills that could be otherwise overlooked.

2 - Official Transcripts and Records; We collect the student's academic background, that is, the subjects studied. These include both the subjects in their field as well as any other general courses they may have taken before graduation. These are obtained directly from the students rather than asking the respective educational institutions. This helps to add to our understanding of what learning choices have been made to that point in time; Students may individually have enrolled in courses that are not required for their major. Those courses would be part of their learning path. In the case of e-Learning to progress to the next course, the current course must be passed. However, several courses are taken concurrently in the case of the transcript. This is why in the current case; grade or performance evaluations didn't have to be used.

3 - Log Data; According to (Rotelli et al., 2021) several cases must be considered while processing log data. In addition, there is a different level of detail that can be acquired. Therefore, restricting the scale of data collected becomes necessary; Here we collect two types of data - the student ID, and subject titles.

E-Learning is becoming an important factor of SSA education. Commonly, producing a Learning path would require having a large sample of e-Learning users' data. If that large sample is unavailable, the workforce is. In an environment of high cultural diversity, identifying the local needs is critical to making adaptable learning tools.

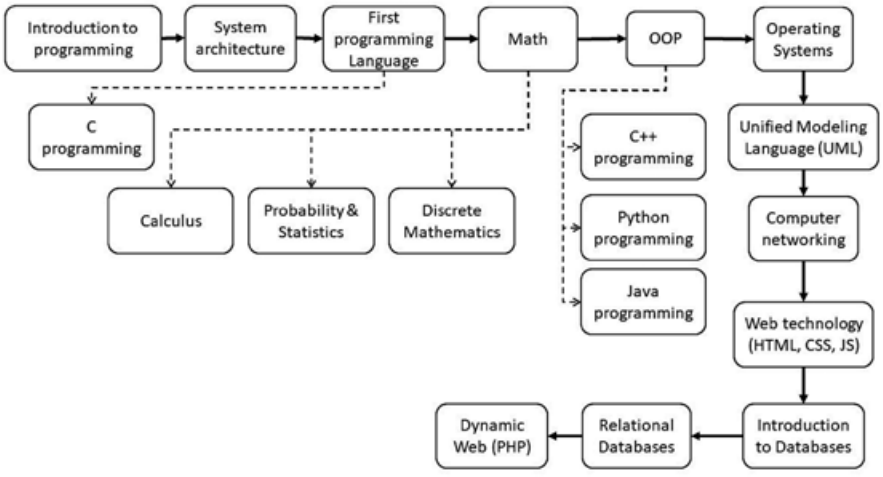

Figure 1: Default learning path

Fig. 1 shows an example of a default learning path containing the most commonly studied subjects in computer science. Dashed arrows show relations between a subject and the more concrete subjects. From the scientific viewpoint, system architecture, introduction to programming, and operating systems are fundamental and necessary knowledge to understand the workings of computers. In the learning on software development, the goal becomes to understand the practical application. Such learning starts the first programming language (FPL) by providing the introduction of programming by C language in the case of Fig.1. In addition, this learning path order is not absolute but it may be changed by the curriculum policies of universities and e-Learning service providers. For example, if the operating system is regarded as a more important subject than FPL on the curriculum policy, it may be moved before FPL.

V. Results and discussion

Our first step was to collect the learning history from four technology graduates in SSA by a survey. We gathered data on what professionals learned during school and between graduation and the end of their first year of employment. The interviewees were an AI engineer from Guinea, a database administrator from Cameroon, a data coordinator from Senegal, and a software engineer from Senegal. Every individual has a different learning path in regard to IT. Table 1 shows examples of data collected from former students in computer science in Senegal. These students have different specializations in IT after graduation.

Table 1: Example of collected data

Status	Skills	Category	Source
Data coordinator	Video editing	Unknown	In house
Software engineer	Computer architecture	Introductory	Transcript
Software engineer	C Programming	Programming class	In school
Database administrator	MERISE modeling	Unknown	In school
Data coordinator	Web cartography	Unknown	In-house

With this information including data shown in Table 1, we could identify a case of skill mismatch as video editing is not part of the main computer science learning path. MERISE is a modeling methodology for software engineering (Avison, 1991). As a subject, it is less common than the Unified Modeling Language. Many young professionals had to undergo training and acquire extra skills during their school-to-work transition, to better adapt to their new environment. This highlights the fact that what is currently being taught in a subject at school is not adequate for real-world survival, in other words, skill mismatch. By relying on professionals' experience and skills, future skill mismatch can be minimized and the need for retraining reduced.

Categorizing the courses allows identifying the course that is missing from the default learning path. They can be reviewed for identification of the ones that are related to IT such as MERISE. Such a method can be used to expand the default learning path. The others would not be part of

an IT learning path.

It is important to note that skills and subjects are not mutually exclusive. In other words, the outcome of studying a subject is often a skill. However not all skills will be taught as a subject, some will form after being repeatedly solicited through the studies of various subjects. From the viewpoint of e-Learning, subjects will be considered. Which subject must be studied to acquire a specific set of skills? While from the viewpoint of skill mismatch skills are being considered. What skills are required to meet specific market demands or needs?

Our goal is to simplify, as much as possible, one's learning path by personalizing it through the addition or deletion of selective content. This might include potentially adding skills that appear unrelated to subjects, like debugging shown in Table 1. Debugging is not often what comes to mind when someone mentions being good at critical thinking. While debugging is certainly an important skill, and one that can be used as a foundation for other skills later, the knowledge of that ability would not have been discovered exclusively through log data generally obtained. That is that such log data would not show that debugging is a skill possibly needed for critical thinking. Using the survey, professionals can express their learning history in terms of skills learned. Based on this history, undiscovered or unexpressed learning data can be obtained.

While having a comprehensive history will help in learning path generation. However, it might be impossible to generate a learning path, unless one can provide a full picture of her/his learning history to that point. This data can be complemented by using the information in resumes and curriculum vitae to give an overall picture of individual learning history. In another viewpoint, the data generated by these sources may prove to be valuable to the content providers to see if the learning path generated through this data can affect the conception of the curriculum (Stabback, 2016).

Our proposed method has some disadvantages. The first one is missing memory in the acquisition process of First-Hand Knowledge. We do not always remember all the details of our learning history. In general, only the best courses are remembered. To reduce the influence of this disadvantage we can use cross-referencing information from people that have studied in the same institution and time. Such cross-referencing information can fill the missing memory. The second disadvantage is hidden information on Official Transcripts and Records. For example, if a student audits a class, it will not be reflected on her/his transcript. Asking for such hidden information explicitly is a simple and effective way to avoid this disadvantage. The third disadvantage is Log Data recorded on plural e-Learning platforms. Generally, Log Data for a user is recorded on a specific e-Learning platform. However, some people may use more than one e-Learning platform. Several tools to visualize or analyze Log Data are being developed as not all content providers can use such data directly (Jugo et al., 2014; Rotelli et al., 2021). Integrating such Log Data remains in future work.

Ⅵ. Conclusion

This paper described a data collection method for generating useful learning paths for young people in SSA. To improve the quality of education in SSA and assist in reaching SDG 4,

generating a personalized learning path from locally provided data is vitally important. Using multiple and diverse local sources further helps to ensure the paths' relevancy. Consulting recent graduates and young professionals proves to be a viable source of information and provides unique insight into skill acquisition, something that is missed through official transcripts and Log Data alone. This First-Hand Knowledge data collection has the potential to not only create highly culturally specific learning paths but also reduce skill mismatch drastically, leading to a more educated and productive workforce in sub-Saharan Africa.

Ⅶ. References

Adelakun, N. O. & Omolola, S. A. (2020). A Pragmatic Study on E-Learning System for Higher Education in Developing Countries, *Proceedings of the 1st National Conference, Nigerian Society of Engineers, Ilaro Branch*, 14-20.
https://dx.doi.org/10.2139/ssrn.3856391

Avison, D. (1991). "MERISE: A European Methodology for Developing Information Systems", *European Journal of Information Systems,* p. 183-191
https://doi.org/10.1057/ejis.1991.33

Balogun, N. A. & Ahlan, A.R. (2015). Issues of e-Learning in developing countries: A Nigerian perspective, 1-3.
https://doi.org/10.1109/ICT4M.2014.7020626

EU Commission (Skill panorama), (2015). "skill mismatch," Last accessed the 10th of June 2020
https://unevoc.unesco.org/home/TVETipedia+Glossary/filt=all/id=568 (access 22-02-05)

GSM Association, (2021). The Mobile Economy Sub-Saharan Africa 2021, 12-15.

Jih, H. J. (1996). The impact of learner's pathways on learning performance in multimedia Computer Aided Learning. *Journal of Network and Computer Applications,* 19, 367–380.
https://doi.org/10.1006/jnca.1996.0025

Jugo, I. & Kovacic, B. & Slavuj, V. (2014). Using Data Mining for Learning Path Recommendation and Visualization in an Intelligent Tutoring System., *MIPRO 2014*, 925-927.
http://dx.doi.org/10.1109/MIPRO.2014.6859700

Kituyi, G. & Tusubira, I. (2013). A framework for the integration of e-Learning in higher education institutions in developing countries. *International Journal of Education and Development using Information and Communication Technology,* 9 (2), 19-36.
http://ijedict.dec.uwi.edu//viewarticle.php?id=1623

Kotoua, I. K. (2015) "The Growing of Online Education in Sub Saharan Africa: Case Study Ghana", *Procedia - Social and Behavioral Sciences,* 191, 2406-2411.
https://doi.org/10.1016/j.sbspro.2015.04.670

Moodle Documentation, (2020). Logs https://docs.moodle.org/311/en/Logs (access 22-02-05)

Rotelli, D. & Fiorentino, G. & Monreale, A. (2021). Making sense of learning log data, arXiv:2106.11071 [cs.CY], 3-11.
https://doi.org/10.48550/arXiv.2106.11071

SDG Center for Africa and Sustainable Development Solutions Network (2019). Africa SDG Index and Dashboards Report 2019. Kigali and New York: SDG Center for Africa and Sustainable Development Solutions Network, 24-31.

Stabback, P. (2016) "What makes a quality curriculum?", B.I.E, 36-38 *in French*