A Curriculum for An Interdisciplinary Program in Digital Humanities

OHYA Kazushi
Tsurumi University
2011-09-13

1 Introduction

Those who are engaged in Digital Humanities are expected to complete studies both in humanities and scientific fields, e.g. computer science or human science. However, it is difficult to expect such training from students, because universities usually do not provide such wide-ranging independent programs for Digital Humanities leading to a college degree. If we hope to cultivate a community of Digital Humanities, we must improve this situation and increase the number of students who have interest in and are engaged in Digital Humanities.

In this paper, first, we will consider the possibility of creating major and minor programs in Digital Humanities based on the case of Tsurumi University, and then show the implausibility. Second, we will confirm the need for a transitional program, or DH Core, as well as a model curriculum for students to use when applying to universities that do not provide it as an off-the-shelf program. Finally, it will be proposed that a) we make a list of potential subfields in order to establish an interdisciplinary program in Digital Humanities with DH core, b) the curriculum for each subfield must be defined by the respective research communities of the fields, and c) we probably need to re-confirm objectives of Digital Humanities.

2 The Case of Tsurumi Univesity

A field of Digital Humanities(hereinafter DH) and other related fields have been taught at Tsurumi University since 2005. We provide four courses specific to DH and some basic courses relating to it. The content in the four courses has changed every year because of two reasons. One is that we would like to experiment with new curricula and methods on students. Another is that we cannot be confident about our methods and curricula that we have taken in all the previous classes. Through the experiment, we have the impression that our curriculum seems to be inappropriate for undergraduate students to learn DH, because the four courses specific to DH are insufficient in quantity, which means that they do not fulfill knowledge areas potentially expected for DH.

Table 1 is a list of selected courses around DH in Tsurumi University. [1]

No.	Year	Sem.[2]	Title
1	1	1	Introduction to Computer Science
2	1	1	Introduction to Library Science
3	1	2	Introduction to Bibliography
4	1	2	Computer Architecture
5	1	2	Database
6	1	2	Networking
7	2	1	Introductory Programming
8	2	1	Discrete Mathematics 1
9	2	1	HTML
10	2	2	DH 1
11	3	1	Discrete Mathematics 2
12	3	1	DH 2-1
13	3	2	DH 2-2
14	3	2	DH 3
15	3	2	Discrete Mathematics 3

Table 1: DH courses in Tsurumi Univ.

[1] The DH 2-1 and 2-2 are provided as a one-year, or two semester course.
[2] In Japan the 1st and 2nd semester are Spring/Summer and Autumn/Winter courses respectively.

In total, fifteen courses around DH are provided in an undergraduate program.[3] The fifteen is not a small number for a minor, or perhaps major program. However, we have felt that there are missing courses for nurturing the embodiment of DH.

2.1 DH 1

Before taking the DH 1, students have learned a way of writing web pages in HTML. Thus, we set the first objective of this course to free students from limited vocabularies of HTML, because students are dominated by the idea that selecting of tags is a substantial activity for encoding texts. When encoding texts, students apt to try to find out layout features and a tricky way of tag combination to realize the appearance in an HTML way. In terms of text encoding, an encoding manner in HTML is 1) to be done under limited circumstance of vocabularies, which means 2) simply a selection of tags from the list, and 3) likely to find out a complicated combination of tags for representing a layout.[4] We believe that encoding in markup languages is a creative process to reflect an intellectual activity in the human brain. If so, encoded texts is possible to contain valuable traces for observing intellectual activity in our brain. To reflect such intellectual activity on encoded texts, the texts should be encoded free from any prefixed manner. It does not mean to exclude a practical approach to share encoded texts according to specifications. In our understanding, requirements for students to learn at the first stage of DH courses are not memorizing prefixed vocabularies but acquiring an inventive stance on encoding texts. Following this premise, we let students markup texts without using HTML tag sets.[5]

In this syllabus, students seem to be disturbed by freedom and at a loss as to what to markup in the beginning. At least two classes seem to be needed for students to get accustomed to be in this situation. But, students finally succeed in finding out existence of some categories in these free-chosen-tags: a type of complementary tag[6] and a type of universal tag.[7] We have the impression that this inductive approach contributes to nurture students' ability of observing characteristics of encoding features, and this development contributes for students to understand a text/document structure. And, we also have the impression that teaching specification about encoding texts at this course is not appropriate in pedagogy. Thus, specifications on encoding texts, e.g. TEI, DocBook, Office Open XML, etc. are excluded from DH 1.[8]

Details in learning are as follows.

	theme	content	class
1	A brief history of DH	concordance, corpus, annotated text.	over 2
2	A practice of encoding	free tagging and collecting features.	over 5 depending on course materials
3	XML Basic	history, grammar, DTD.	over 3
4	XML family and processing	a brief introduction	1

Table 2: DH 1

2.2 DH 2-1

An objective of DH 2-1 is to let students learn actual problems in encoding Japanese literature and a way of the management. In the case of encoding Japanese literature, character handling is an inevitable obstacle that requires us to decide a fundamental attitude to a way of encoding texts when doing the first letter. It is not so easy to move an encoding stage from characters to words. We need to take a certain amount of time for students to consider problems in encoding characters. As a way of this transition, we adopt a group work or a workshop style for students to find problems and decide an encoding stance or philosophy. Therefore, since attendees are different, an encoding style shared in a class changes every year.

In our practice, we use a serial novel in a daily newspaper published about 100 years ago. There are four original versions: manuscripts, Tokyo newspaper version, Osaka newspaper version, and the first book version. Through encoding and collating these four versions, students learn many ways of annotation in a digital form,

[3] There are other courses for library science, bibliography, and information science leading to a bachelor degree in Department of Library, Archival, and Information Studies, Tsurumi University, Japan.
[4] At this stage we do not teach a method to specify a layout, e.g. CSS and XSLT.
[5] Then, we teach a document structure and a way to specify.
[6] e.g. chapters and pages.
[7] e.g. emphasis and rubies.
[8] This stance might be different from one adopted in other many workshops and summer schools.

a way of collation, and a variety of editing policies. At a final stage, we introduce specifications for encoding style, mainly TEI.[9] In this course, however, we do not go into details of TEI.

Details in learning are in table 3.

	theme	content	class
1	Encoding and Editing	a type of editing and a final objective of encoding	over 2
2	Encoding and Processing	a type of processing of text data and the practice	over 1
3	Encoding and Publishing	a user interface matching with the objective	over 1
4	Markup Languages	history, type, theory, XML	over 1
5	Practice of encoding Japanese literature	encoding Japanese literature	over 4
6	Collation in a Digital Form	collating encoded texts	over 3
7	Specifications: TEI	an introduction to TEI	1

Table 3: DH 2-1

2.3 DH 2-2

An objective of DH 2-2 is to let students learn XML data processing with specifications of the XML family, e.g. XPath, XSLT and XQuery. As an editing tool we use JEdit[6] because 1) it supports Japanese texts well, 2) it supports many plug-ins, 3) it is easy to install plug-ins, 4) it is lightweight, 5) it can support a keymap in Emacs style,[10], and 6) it is available free of charge.

In this course a troublesome learning topic is XSLT, because 1) there is no free good engine and commercial products are expensive for universities,[11] 2) XSLT is an XML application specific to functional languages,[12] and 3)in most cases, it is enough to use XQuery both in transforming and retrieving from XML data. On the other hand, we have the impression that XQuery is very instructive and useful for students, because XQuery is 1) easy to learn for students who are accustomed to use procedural programming languages, 2) able to use step-by-step procedures by using a 'let' sentence, 3) supported by high speed engines built in native XML DBs, and 4) adopted in DB systems that students use for their own activities including graduation theses. We would like to use JavaScript in teaching CSS and DOM if it is possible.

Details in learning are in table 4.[13]

	theme	content	class
1	CSS	HTML and CSS1 and CSS2	over 2
2	XPath	syntax, practice on JEdit	over 1
3	DOM	concept, sytax, and processing	over 1
4	XSLT	concept, syntax, practice on JEdit	over 2
5	XQuery	concept, syntax, practice on JEdit	over 4
6	XML DB	RDB and XMLDB, installing Berkeley DB XML, practice	over 2

Table 4: DH 2-2

2.4 DH 3

DH 3 is a course for encoding a large map in image data which is made through being captured by digital cameras and composed by Photoshop. Making a large image of multiple images taken by cameras is a very tedious and difficult process because of multiple distortions, e.g. optical and chromatic aberration. To avoid these kinds of changes, we can resort to scanners instead of cameras. However, scanners cannot be used in the case that 1) the map is too large to be set on a table, 2) the map is not permitted to be folded, 3) the map is sometimes not permitted to touch directly a rigid material like steel or glass, and 3) waved materials on which especially characters sink into a groove are inappropriate for the subjects of scanners. Therefore, we adopt

[9] We introduce TEI in order to parse and validate annotated texts with TEI's DTD.

[10] Emacs is used also in other courses.

[11] This may be peculiar to Japan.

[12] XSLT is the first functional language for many students.

[13] As you can see, the schedule is relatively condensed.

digital cameras as a device for digitization of large materials. Unfortunately we do not teach a way of recording metadata on images in this course due to time constraints. As a publication system we use Zoomify Express[7].

Details in learning are in table 5.

	theme	content	class
1	Introduction to Image encoding	a way of digital conversion, a type of format, quality control, metadata, an access to digital image	1
2	Light Sensing	a mechanism of camera, light intensity, lens, shutter speed, diaphragm	1
3	Practice of photo taking	practice	2
4	Encoding a large map	taking photos of a map	over 3
5	Editing large image	combining multiple images into one image on Adobe Photoshop	over 4
6	Publication	storing a large image into a publication system with Zoomify.	1

Table 5: DH 3

2.5 Successes and Challenges

We have the impression that students succeed in acquiring sound encoding philosophy and mastering methods of encoding materials. However, we have felt a few shortcomings in our courses as follows.

- We do not teach specifications about encoding texts.
- We do not teach data processing independently from XML software.
- We do not teach a stochastic approach to encoded texts.[14]

If it is possible to expand our program, we would like to add new courses as such. However, it is also true that the more we provide the number of DH courses, the fewer chances for students to study humanities and other fields. If we increase the number of courses to master computer science, it certainly restricts students' chances to take courses in humanities. Actually, we feel inadequacy of students' achievement in humanities. In our case, the number of courses of computer science affects students' activitis in humanities.[15] The number of our present courses, at least fifteen, is not short as a minor program, however, we face on problems caused by a shortage of courses. We are in the dilemma: the more courses we provide, the less chances students learn. In order to seek a solution to this problem, as a preliminary, we make a list of ideal courses that are required for DH without any restriction.

3 DH Core

Many researchers engaged in DH could agree that DH is based on both humanities and scientific fields such as computer science and human science, and also that DH cannot be made just of the both knowledge according to a compositional principle. We believe many DH researchers agree that there exist independent courses specific to DH. In this paper, we call such a set of independent courses DH Core. As such DH Core, we suppose the curriculum in table 6.[16]

In addition to this curriculum of elementary courses, advanced courses in each field would be needed. As a result, the total number of courses in DH core might be around 15.[17]

[14] In a class for discrete mathematics we teach a foundation of probability and statistics.

[15] In our university there is the upper limit of the course that students can take within a year. Students can earn at most 44 credits per year.

[16] We understand that this list must not be a general model agreed by every researcher in DH. The point is that DH community need to develop a list of core courses for DH.

[17] If a second introductory course like generally-called DH 2 is needed, there are 16 courses in total.

	course	theme
1	Introduction to DH	history, annotated text, methods, objective.
2	XML Core	HTML/CSS, markup language, XPath/XPointer, DTD/XML Schema, Infoset/DOM, encoding model
3	XML Applications	XML families
4	XML Data Processing	utility/editor, XSLT, XQuery, script language, Java
5	Text Encoding	data structure in texts, data structure of encoded texts, characters, features beyond text, specifications
6	Non-Text Encoding	still images, motion picture, sound, kinesiology, printing, music, stage
7	Publication and Preservation	bibliography, metadata, web service, HTML 5, ePUB, XML DB, link management
8	Encoding Management	editing policy, documentation, quality control, resource control, collation

Table 6: DH Core

4 A Possibility of Virtual Independent Program

4.1 Premises

DH is an interdisciplinary field, which typically consist of fields of humanities and STEM such as computer science, computer engineering, or human science. DH requires students to take double majors or more. Table 7 and 8 are sample lists of computer science and linguistics with recommendation for DH respectively.[18]

	course	for DH
1	CS1(a first introductory course)	✓
2	CS2(a second introductory course)	✓
3	Hardware	
4	Data Structure and Algorithm	✓
5	Design of Human Interface	✓
6	Information System	✓
7	Computer Graphics	
8	Computer Network	✓
9	Programming Language	✓
10	Visualization	
11	Operating System	
12	Software Engineering	
13	Compiler	
14	Formal Language	✓
15	Discrete Mathematics	✓
16	Artificial Intelligence	✓
17	Natural Language Processing	✓
18	Multimedia Handling	
19	Web System	✓

Table 7: Computer Science Courses

	course	for DH
1	Introduction to LNG	✓
2	Phonetics	✓
3	Phonology	✓
4	Syntax	✓
5	Morphology	✓
6	Semantics	✓
7	Typology	✓
8	Socio-Linguistics	
9	Field Linguistics	✓
10	Psycho-Linguistics	✓
11	Language Education	✓
12	Computational Linguistics	✓
13	Formal Language	
14	Individual Languages	✓

Table 8: Linguistics Courses

This kind of lists can be defined only by experts in each field. Students can consult these lists for pursuing DH by themselves. However, unfortunately, related communities do not provide such list. Each expert now engaged in DH should provide a recommendation list of knowledge areas or sample courses for DH. The list must be helpful for eager students in DH.

[18] There is no specific reason to select linguistics and computer science as representatives of each field.

In summing up, the number of presumable courses, a total of 39 is expected for a DH program in this example.

4.2 Provisional Calculations: around 30

Each country and university adopts a different education system. But, usually credits and hours are used as a unit for calculation in the case of mutual recognition of degree approval. For example, the required credits for graduation in US and Japan are usually around 130 and 120 respectively. And, it means that in theory there are 40 and 65 courses on average in US and Japan respectively. However, the point here is an education system for profession. In an international degree exchange system for engineering,[19] 1600 hours are basic criterion to approve the total education, and among them 900 hours are for a specific field of knowledge. In the case of Japan, 1600 hours divided by 130 credits is 12.3 hr/credit, then 900 hr means about 73 credits. Since one course basically earns 2 credits, then 36.5 courses are for education of a specific field. In the case of US, 1600 hours divided by 120 credits that is required for graduation is 13.3 hour/credit, which means 67.5 credits are required for graduation. Since one course basically earns 3 credits, then about 22.5 courses are for a specific field of education(900hr required). All these numbers include credits for graduation thesis. Then, the total number of courses is under 36.5 and 22.5 respectively. Of course, the number of credits is assigned to each course depends on universities and areas. However, as an estimate, even if we set the target number of courses around 30 as the mean of 22.5 to 36.5, the number of courses in an ideal DH program surpasses this target number and the range.[20]

4.3 Architectures of DH Program

According to the estimation so far, DH is hard to be set as an independent undergraduate program in an existing university system, because 1) ideally, more than two sure knowledge fields are required for DH, and 2) a total of 39 courses in undergraduate programs exceeds the expectations for students' studies. Therefore, we need to seek a way to absorb and reduce the number of courses in an undergraduate program. A graduate program is a prime alternative for such an overflow of courses in the undergraduate program. But, it is not a good policy to set all of the courses only in graduate programs, because DH is still not an appealing field. DH cannot be attractive to students who have mastered traditional academic areas that are steady to make a career. If possible, it is better to provide DH courses in both undergraduate and graduate programs. The DH courses in undergraduate programs would be a prelude to fostering young researchers in DH, and those in graduate programs would provide students with advanced knowledge or practices in DH, which would be attractive to students in other fields. In order to realize it, there are at least three problems: 1) whether existing universities allow this kind of program leading to a college degree, 2) what kind of degree is given to the students, and 3) whether such a list of courses can be regarded as an independent program. If we can put aside these problems, table 9 and 10 seem to be architecture of virtual DH courses.[21]

	CS	Ling	DH
Graduate		m	m
Undergraduate	M	m	m

Table 9: Architecture A

	CS	Ling	DH
Graduate		M	m
Undergraduate	M		m

Table 10: Architecture B

In both tables, CS and Ling are exchangeable. Table 9 is the most ideal course architecture, however, it brings another program: Ling, in this case, has to be separated into undergraduate and graduate programs. Architecture B of table 10 requires students to take a major originally held in an undergraduate programs at the graduate program. It is undeniable that these architectures are idealistic plans. Table 11 is a more realistic architecture. However, as we saw previously, it requires students to work hard without high expectations about their career.

An education system of CS is recently changing to be applied to other STEM fields. For example, ACM and IEEE are undertaking setting new curricular guidelines to bridge multi-disciplinary work that can be called "Computational X"[3]. As such a practice, several universities start new interdisciplinary programs, in which CS is in a position of extended minor[1, 2]. In these cases, even though computer science is for teaching

[19] International Engineering Alliance, and, as signatories, JABEE; Japan Accreditation Board for Engineering Education.

[20] It is just a calculation, a play with numbers. However, in order to define a referential model, we need to frame the total number of courses.

[21] Upper letter M and lower letter m represent major and minor courses respectively.

	CS	Ling	DH
Graduate			m
Undergraduate	M	M	m

Table 11: Architecture C

computational methods, at least a five year program is expected because the total number of courses to master a basic theory and practical methods for simulation and programming goes beyond the range of 4 year courses.

DH is not a provider of problem-solving methods but an application made of multiple fields. Thus, more courses are ideally expected rather than those for computational programs. It means that an ideal DH programs would be an over four years program, which means that it requires taking graduate classes.

5 Necessities for Sustainable Research Community

From the above observation, we would like to propose the followings as preparation to realize a DH program.

1. We must build a consensus on courses for DH Core which is used as a referential model for students.
2. We must make a list of potential subfields in order to establish an interdisciplinary program with DH core.
3. A curriculum for each subfield must be defined by the respective research communities of the field.
4. We need to seek a possibility to set a DH program in both undergraduate and graduate schools.
5. We need to reconfirm and share objectives for Digital Humanities within it.

The last point is provocative but crucial. In our education in Tsurumi University, we have felt ourselves failing in indicating clear objectives of encoding literatures, especially when teaching data processing as follows.

- We provide students with only basic examples of processing XML data, and cannot show more practical examples.[22]
- It affects students' motivation for learning.
- It would lead to jeopardizing a raison d'etre for DH in students.

If we choose traditional targets in Humanities without clear objectives suitable for scientific approaches, we would fail to adapt STEM methods or approaches to them, and would get the same old results. It means that we must redefine objectives of DH different from those taken in traditional Humanities. If not, we would fail to configure an interdisciplinary field. DH is an interdisciplinary research field, hence, DH is not Humanities nor one of the STEM field. We cannot achieve objectives that traditional CS seeks by using methods that traditional Humanities adopts, and vice versa. Therefore, for example, we must 1) achieve objectives that do not exist in CS, by using methods that do not exist in Humanities, or 2) achieve objectives that do not exist in Humanities, by using methods that do not exist in CS. The latter is obviously impossible. Thus, in Digital Humanities, we solve problems that do not exist in CS but relating to Humanities, by using methods that do not exist in Humanities but relating to CS.[23]

6 Conclusion

In this paper, we proposed several preparatory activities to strengthen a research community of Digital Humanities based on the case of Tsurumi University. We should recall that a research field of Digital Library has been shrinking because of failure to set a clear objective and a stable education system. Interdisciplinary activities can be seen in many places e.g. eScience, computational science, and so on. We should take these as references for enhancing DH community with humility. Digital Humanities is a new field in an old education system. We need new architecture for education systems, and good objectives and methods for Digital Humanities.

[22] Even if we provide practical examples, they are still "toy data" in a sense.

[23] This is the case where we choose CS as a related field to DH. But, this logical deduction could be applied to the case of other fields.

References

[1]O. Astrachan et.al., 2011, “CS Principles: Piloting a New Course at National Scale”, *SIGCSE 2011*, ACM

[2]J.Caristi et.al., 2011, “Starting a Computational Science Program”, *SIGCSE 2011*, ACM

[3]M.Sahami et.al., 2011, “Setting the Stage for Computing Curricula 2013”, *SIGCSE 2011*, ACM

[4]International Engineering Alliange, http://www.ieagreements.org/

[5]JABEE, http://www.jabee.org/

[6]JEdit, http://www.jedit.org/

[7]Zoomify Express, http://www.zoomify.com/express.htm