Information Retrieval

GLIS 616

Fall 2017

Quick Links

WARNING: this syllabus is provided for informational use only. The specific content and assignments may change before the start of the course. Students of this course should not use this document but instead retrieve the official version that can be downloaded from the course management site.

Instructor: Prof. Charles-Antoine Julien < charles.julien@mcgill.ca >

Duration: Sept. 5th - Nov. 28th, 2017

Class times: Tuesdays, 2:30pm - 5:30pm

Location: LEA 212

Labs: LEA 212; during 2nd half of each class

Office hours : Appointment upon request

Description

Theoretical and applied concepts of information retrieval in a variety of digital environments and in relation to textual and multimedia data.

The course is roughly divided into two parts:

1. **Excel VBA**

   Programming using Excel Visual Basic (VBA) for Applications.

2. **Text parsing, indexing and querying**

   Automatic text processing and develop a functional text indexing and searching system using Excel VBA.

Learning Objectives

- To understand how information retrieval principles are implemented in various digital information environments.
- To acquire working knowledge and experience of the Excel VBA programming language.
- To implement a basic textual information retrieval system using Excel VBA.

Course schedule

Week	Date	Title	Readings
1	Sept.5	Introduction Excel Formulas	Walkenbach, ch. 3
2	Sept.12	Introduction to Excel Visual Basic for Applications (VBA)	Walkenbach, ch. 7
3	Sept.19	VBA 2	Walkenbach, ch. 7-8
4	Sept.26	VBA 3 **Alumni working in IT presentation**	Walkenbach, ch. 8

5	Oct.3	Text File Handling in VBA	Walkenbach, ch. 27
6	Oct.10	Study Break	
7	Oct.17	Textual Information Storage and Access	Manning, C. D., Raghavan, P. & Schütze (2007). H. *Introduction to Information Retrieval.* Cambridge University Press, Chapters 2-3
8	Oct.24	**Mid-Term (25%)** On lab computers All documentation allowed.	Modifications of exercises assigned during weeks 2-5
9	Oct.31	IR Models	G. Salton, A. Wong, and C. S. Yang. 1975. *A vector space model for automatic indexing.* Commun. ACM 18, 11 (November 1975)
10	Nov.7	A complete indexing and querying system. (Final Project Overview)	
11	Nov.14	Web IR	• Page, L., Brin, S., Motwani, R., Winograd, T. *The page rank citation ranking: Bringing order to the web* (1998). Available at http://ilpubs.stanford.edu:8090/422/ • Bar-Ilan, J. (2006). *Web links and search engine ranking: the case of Google and the query 'jew'.* Journal of the American Society for Information Science and Technology 57 (12), 1581-1589
12	Nov.21	**Final Project Part 1** due at 11:59pm Music Information Retrieval Image Retrieval Audio/Video Retrieval	Downie, Stephen J., (2003) *Music information retrieval.* Annual Review of Information Science and Technology.37(1), 295-340
13	Nov.28	Review, Project Lab Session	
14	Dec.5	**Final project (50%)** due at 11:59pm	
14	Dec.7, 9-12pm (LEA 738)	**Final Exam (25%).**	Slides, readings, and skills associated with the final project.

Readings

The following are freely available from McGill libraries.

- Walkenbach, J. (2010). *Excel 2010 power programming with VBA.* Hoboken, N.J: Wiley Pub. http://site.ebrary.com/lib/mcgill/docDetail.action?docID=10381033 (See other sources from http://mcgill.worldcat.org/title/excel-2010-power-programming-with-vba/oclc/647915344&referer=brief_results)

- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval.* New York: Cambridge University Press. http://proquest.safaribooksonline.com.proxy2.library.mcgill.ca/9780511410802?tocview=true

- Individual articles are freely available from McGill Libraries.

Assessments

Excel VBA Exercises

- Excel VBA exercises included in first 5 lectures will not be graded but **you must submit them** the night before the following week's class at 11:59pm.
- Failure to submit any assignment will result in an automatic **25% deduction** on the mid-term exam.

Mid-Term exam (25%, individual, in lab)

- Practical Excel VBA exam performed on lab computers during class.
- All documentation permitted.
- Questions are modifications of the exercises assigned during first 5 lectures.
- You do not have time to learn during the exam; therefore the main difficulty will be time management.

Final Project (50%, individual)

You will build a functional search system using Excel VBA. The system will be developped in two parts.

Part 1: Indexing and term weighting

The objectives of part 1 are the following:

1. Read a collection of text files provided on myCourses;
2. Create the dictionary and postings of the collection;
3. Assign weights to each dictionary term based on VSM tf-idf weigthing.

Part 1 Instructions:

- Part 1 will be submitted as a single Excel file (.xlsm).
- Part 1 will not be graded specifically but it will serve as a feedback point to ensure completion of the final project.
- Failure to submit part 1 will result in an automatic **20% deduction** on the complete final project grade.

Part 2: Complete Query System

Using the results of part 1, the complete system will:

1. Accept a user query of one word and return the matching documents;
2. Accept a user query of one word and rerturn the matching documents **ranked using tf-idf**;
3. Perform the previous two requirements for multi-word queries.

Final Exam (25%, individual, in class)

- The final exam covers theory (i.e., slides and readings) and your practical knowledge of Excel VBA.
- **Note:** content needed to complete the final project is covered by the final exam.
- This is a traditional paper based exam performed in the classroom during regular class time.
- All documentation is allowed up to 4 double sided 8.5 by 11 in. paper sheets. You can use both sides of the paper for a total of 8 pages of notes. Notes can be in any format (e.g., printed, handwritten, etc.).

Submitting your assignments

- Your assignment must be submitted electronically using the Assignments tool in myCourses.
- Multi file submissions, if any, must be packaged as a single zip file. No other format is accepted (e.g., rar).
- Late assignments will not be accepted, and will receive a grade of zero (0). Extensions are only granted in the most exceptional of circumstances, and they must be requested at least one week before the assignment is due.

General Information

Academic Integrity

McGill University values [academic integrity](#). Therefore, all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Council and Disciplinary Procedures. See www.mcgill.ca/students/srr/honest/ for more information.

Academic Policies

Additional policies governing academic issues which affect students can be found in the [McGill Charter of Students' Rights.](#)

Language of Instruction

In accord with McGill University's Charter of Students' Rights, students in this course have the right to submit in English or in French any written work that is to be graded.

Students with disabilities

If you have a disability please consult the [Office for Students with Disabilities](#)

Syllabus may change

In the event of extraordinary circumstances beyond the University's control, the content and/or evaluation scheme in this course is subject to change.