Project 1. MIPS Assembler

Due 23:59, April. 1st

TA: Seunghyo Kang, Daehyeon Baek

1. Introduction

The objective of the first project is to implement a **MIPS ISA assembler**. The assembler is a tool which converts assembly codes to a binary file. This project is intended to help you understand the MIPS ISA instruction set.

The assembler you are going to design is a simplified assembler which does not support linking process, and thus you do **not need to add the symbol and relocation tables** for each.

You should implement the assembler which can convert a subset of the instruction set shown in the following table. In addition, your assembler must handle labels for jump/branch targets, and labels for the static data section.

**If you have any questions related to the project, please ask them on the email(<u>cs311 ta@casys.kaist.ac.kr</u>) or office hour. The assigned TA(Seunghyo Kang, Daehyeon Baek) will answer them whenever possible.

- Instruction Set

The detailed information regarding instructions are in the green card page of textbook They are also attached in the following two pages.

ADDIU	ADDU	AND	ANDI	BEQ	BNE	J
JAL	JR	LUI	LW	LA*	NOR	OR
ORI	SLTIU	SLTU	SLL	SRL	SW	SUBU

MIPS Reference Data

	-		
A	V	Ï	A
		F	
	A	1	y
A		-	

1

CORE INSTRUCTION	ON SE	Т			OPCODE
		FOR-			/ FUNCT
NAME, MNEMO	NIC	MAT	The state of the s		(Hex)
Add	add	R	R[rd] = R[rs] + R[rt]	(1)	0 / 20 _{hex}
Add Immediate	addi	I	R[rt] = R[rs] + SignExtImm	(1,2)	8 _{hex}
Add Imm. Unsigned	addiu	I	R[rt] = R[rs] + SignExtImm	(2)	9 _{hex}
Add Unsigned	addu	R	R[rd] = R[rs] + R[rt]		0 / 21 _{hex}
And	and	R	R[rd] = R[rs] & R[rt]		0 / 24 _{hex}
And Immediate	andi	I	R[rt] = R[rs] & ZeroExtImm	(3)	c _{hex}
Branch On Equal	beq	I	if(R[rs]==R[rt]) PC=PC+4+BranchAddr	(4)	4 _{hex}
Branch On Not Equal	bne	I	if(R[rs]!=R[rt]) PC=PC+4+BranchAddr	(4)	5 _{hex}
Jump	j	J	PC=JumpAddr	(5)	2 _{hex}
Jump And Link	jal	J	R[31]=PC+8;PC=JumpAddr	(5)	3 _{hex}
Jump Register	jr	R	PC=R[rs]		0 / 08 _{hex}
Load Byte Unsigned	1bu	1	R[rt]={24'b0,M[R[rs] +SignExtImm](7:0)}	(2)	24 _{hex}
Load Halfword Unsigned	lhu	I	R[rt]={16'b0,M[R[rs] +SignExtImm](15:0)}	(2)	25 _{hex}
Load Linked	11	I	R[rt] = M[R[rs] + SignExtImm]	(2,7)	30 _{hex}
Load Upper Imm.	lui	I	$R[rt] = \{imm, 16'b0\}$		f _{hex}
Load Word	lw	I	R[rt] = M[R[rs] + SignExtImm]	(2)	23 _{hex}
Nor	nor	R	$R[rd] = \sim (R[rs] \mid R[rt])$		0 / 27 _{hex}
Or	or	R	R[rd] = R[rs] R[rt]		0 / 25 _{hex}
Or Immediate	ori	I	R[rt] = R[rs] ZeroExtImm	(3)	d _{hex}
Set Less Than	slt	R	R[rd] = (R[rs] < R[rt]) ? 1 : 0		0 / 2a _{hex}
Set Less Than Imm.	slti	I	R[rt] = (R[rs] < SignExtImm)? 1	: 0(2)	a _{hex}
Set Less Than Imm. Unsigned	sltiu		R[rt] = (R[rs] < SignExtImm) $? 1: 0$	(2,6)	b _{hex}
Set Less Than Unsig.	sltn	R	R[rd] = (R[rs] < R[rt]) ? 1 : 0		0 / 2b _{hex}
Shift Left Logical	sll	R	$R[rd] = R[rt] \ll shamt$	(-)	0 / 00 _{hex}
Shift Right Logical	srl	R	R[rd] = R[rt] >>> shamt		0 / 02 _{hex}
Shirt Right Logical	211		M[R[rs]+SignExtImm](7:0) =		
Store Byte	sb	I	R[rt](7:0)	(2)	28 _{hex}
Store Conditional	sc	Ι	M[R[rs]+SignExtImm] = R[rt]; R[rt] = (atomic) ? 1 : 0	(2,7)	38 _{hex}
Store Halfword	sh	Ι	M[R[rs]+SignExtImm](15:0) = R[rt](15:0)	(2)	29 _{hex}
Store Word	SW	I	M[R[rs]+SignExtImm] = R[rt]	(2)	
Subtract	sub	R	R[rd] = R[rs] - R[rt]	(1)	0 / 22 _{hex}
Subtract Unsigned	subu	R	R[rd] = R[rs] - R[rt]		0 / 23 _{hex}
	(2) Sig (3) Ze (4) Br	gnExtl roExtl anchA	se overflow exception from = { 16{immediate[15]}, immediate from = { 16{1b'0}, immediate } from = { 14{immediate[15]}, immediate from the diagram of	ediate,	
			s considered unsigned numbers (vi est&set pair; R[rt] = 1 if pair atom		
DACIO INCTRIICT	ONE			,	

BASIC INSTRUCTION FORMATS

R	opcode		rs		rt		rd	shamt	funct
10	1	26		21 20		16 15	2000		5 0
I	opcode		rs		rt		HEI	immediate	e
	31	26	25	21 20		16 15			0
J	opcode						address		
	31	26	25	7/1			THE STATE OF		n

ARITHMETIC CORE INSTRUCTION SET

7011711112110 00112			©	/ FMT /FT
	F	OR-		/ FUNCT
NAME, MNEMONI		MAT		(Hex)
Branch On FP True bo		FI	if(FPcond)PC=PC+4+BranchAddr (4)	The second second
Branch On FP False bo	clf	FI	if(!FPcond)PC=PC+4+BranchAddr(4)	11/8/0/
Divide d	liv	R	Lo=R[rs]/R[rt]; Hi=R[rs]%R[rt]	0//-1a
Divide Unsigned di	ivu	R	Lo=R[rs]/R[rt]; Hi=R[rs]%R[rt] (6)	0///1b
FP Add Single ad	id.s	FR	F[fd] = F[fs] + F[ft]	11/10//0
FP Add	id.d	FR	${F[fd],F[fd+1]} = {F[fs],F[fs+1]} +$	11/11//0
Double	ia.a	LK	{F[ft],F[ft+1]}	11/11//0
FP Compare Single co	x.s*	FR	FPcond = (F[fs] op F[ft])?1:0	11/10//y
FP Compare	x.d*	FR	$FPcond = (\{F[fs], F[fs+1]\} op$	11/11//v
Double			{F[ft],F[ft+1]})?1:0	11/11//
			=, <, or <=) (y is 32, 3c, or 3e)	**********
And the state of t	V.S	FR	F[fd] = F[fs] / F[ft]	11/10//3
FP Divide	v.d	FR	${F[fd],F[fd+1]} = {F[fs],F[fs+1]} /$	11/11//3
Double		ED	{F[ft],F[ft+1]}	11/10/ /2
	ıl.s	FR	F[fd] = F[fs] * F[ft]	11/10//2
FP Multiply mu	ıl.d	FR	${F[fd],F[fd+1]} = {F[fs],F[fs+1]} *$	11/11//2
Double		ED	{F[ft],F[ft+1]}	11/10//1
	ıb.s	FR	F[fd]=F[fs] - F[ft]	11/10//1
FP Subtract Double	ib.d	FR	${F[fd],F[fd+1]} = {F[fs],F[fs+1]} -$	11/11//1
	wc1	I	$\{F[ft],F[ft+1]\}$ $F[rt]=M[P[rt]+SignFut[rtm]] \qquad (2)$	31//
Load FP Single Load FP	MCT	1	F[rt]=M[R[rs]+SignExtImm] (2 F[rt]=M[R[rs]+SignExtImm]; (2	
Double 10	dc1	I	F[rt]=M[R[rs]+SignExtImm]; (2) F[rt+1]=M[R[rs]+SignExtImm+4]	35//
	fhi	R	R[rd] = Hi	0 ///10
	flo	R	R[rd] = H	0///12
	fcO	R	R[rd] = CR[rs]	10 /0//0
	ult	R	$\{Hi,Lo\} = R[rs] * R[rt]$	0///18
* *	iltu	R	$\{Hi,Lo\} = R[rs] * R[rt] $ $\{Hi,Lo\} = R[rs] * R[rt] $ (6)	
	sra	R	$R[rd] = R[rt] \gg \text{shamt}$	0///3
	wc1	I	M[R[rs]+SignExtImm] = F[rt] (2)	
Store FP	WCT.	1	M[R[rs]+SignExtInm] = F[rt]; (2)	
Double	dc1	I	M[R[rs]+SignExtImm+4] = F[rt+1]	3d//
Dodolo			military organization of titter	

(2) OPCODE

FLOATING-POINT INSTRUCTION FORMATS

FR	opcode	fmt	ft	fs	fd	funct
	31 26	25 21	20 16	15 11	10 6	5 0
FI	opcode	fint	ft		immediate	
	31 26	25 21	20 16	15		0

PSEUDOINSTRUCTION SET

NAME	MNEMONIC	OPERATION
Branch Less Than	blt	if(R[rs] < R[rt]) PC = Label
Branch Greater Than		if(R[rs]>R[rt]) PC = Label
Branch Less Than or Equal	ble	$if(R[rs] \le R[rt]) PC = Label$
Branch Greater Than or Equa	d bge	if(R[rs] >= R[rt]) PC = Label
Load Immediate	li	R[rd] = immediate
Move	move	R[rd] = R[rs]

REGISTER NAME, NUMBER, USE, CALL CONVENTION

NAME	NUMBER	USE	PRESERVEDACROSS A CALL?
Szero	0	The Constant Value 0	N.A.
Sat	1	Assembler Temporary	No
\$v0-\$v1	2-3	Values for Function Results and Expression Evaluation	No
\$a0-\$a3	4-7	Arguments	No
\$t0-\$t7	8-15	Temporaries	No
\$s0-\$s7	16-23	Saved Temporaries	Yes
\$t8-\$t9	24-25	Temporaries	No
\$k0-\$k1	26-27	Reserved for OS Kernel	No
Sgp	28	Global Pointer	Yes
Ssp	29	Stack Pointer	Yes
\$fp	30	Frame Pointer	Yes
Sra	31	Return Address	Yes

DPCODES, BASE CONVERSION, ASCII SYMBOLS									
	(1) MIPS	(2) MIPS				ASCII	D	Hexa-	ASCII
opcode	funct	funct	Binary	Deci-	deci-	Char-	Deci-	deci-	Char-
(31:26)	(5:0)	(5:0)		mal	mal	acter	mal	mal	acter
(1)	sll	add.f	00 0000	0	0	NUL	64	40	(a)
		sub.f	00 0001	1	1	SOH	65	41	A
j	srl	mul.f	00 0010	2	2	STX	66	42	В
jal	sra	div.f	00 0011	3	3	ETX	67	43	C
beq	sllv	sqrt.f	00 0100	4	4	EOT	68	44	D
bne	1	abs.f	00 0101	5	5	ENQ	69 70	45 46	E F
blez	srlv	mov.f	00 0110 00 0111	6	7	ACK BEL	71	47	G
bgtz addi	srav jr	neg.f	00 1000	8	8	BS	72	48	H
addiu	jalr		00 1001	9	9	HT	73	49	Ī
slti	movz		00 1010	10	a	LF	74	4a	J
sltiu	movn		00 1011	11	b	VT	75	4b	K
andi	syscall	round.w.f	00 1100	12	С	FF	76	4c	L
ori	break	trunc.w.f	00 1101	13	d	CR	77	4d	M
xori		ceil.w.f	00 1110	14	e	SO	78	4e	N
lui	sync	floor.w.f	00 1111	15	f	SI	79	4f	0
(2)	mfhi		01 0000	16	10	DLE	80	50	P
(2)	mthi mflo	marin f	01 0001	17	11 12	DC1 DC2	81	51 52	Q R
	milo	movz.f	01 0010	19	13	DC3	83	53	S
	INCLO	movii.j	01 0100	20	14	DC4	84	54	T
			01 0101	21	15	NAK	85	55	U
			01 0110	22	16	SYN	86	56	V
			01 0111	23	17	ETB	87	57	W
	mult		01 1000	24	18	CAN	88	58	X
	multu		01 1001	25	19	EM	89	59	Y
	div		01 1010	26	la	SUB	90	5a	Z
	divu		01 1011	27	16	ESC	91	5b]
			01 1100	28	10	FS	92 93	5c 5d	1
			01 1101 01 1110	29 30	1d 1e	GS RS	93	5e	,
			01 1111	31	1f	US	95	5f	
lb	add	cvt.s.f	10 0000	32	20	Space	96	60	
lh	addu	cvt.d.f	10 0001	33	21	!	97	61	a
lwl	sub		10 0010	34	22	11	98	62	b
lw	subu		10 0011	35	23	#	99	63	c
1bu	and	cvt.w.f	10 0100	36	24	\$	100	64	d
lhu	or		10 0101	37	25	%	101	65	e
lwr	xor		10 0110		26	&	102	66	f
	nor		10 0111	39	27		103	67	g
sb			10 1000		28	(104	68 69	h
sh swl	el+		10 1001 10 1010	41 42	29 2a) *	105	6a	j
SWI	slt sltu		10 1010	43	2b	+	107	6b	k
OW	9100		10 1100		2c	,	108	6c	1
			10 1101	45	2d	-	109	6d	m
swr			10 1110		2e		110	6e	n
cache			10 1111	47	2f		111	6f	0
11	tge	c.f.f	11 0000		30	0	112	70	р
lwcl	tgeu	c.un.f	11 0001	49	31	1	113	71	q
lwc2	tlt	c.eq.f	11 0010		32	2	114	72	r
pref	tltu	c.ueq.f	11 0011	51	33	3	115	73	S
1 -1 -1	teq	c.olt.f	11 0100		34	5	116	74 75	t
ldc1	tne	c.ult.f	11 0101 11 0110		35 36	6	117	76	u v
ldc2	tne	c.ole.f	11 0111	55	37	7	119	77	W
SC		c.sf.f	11 1000		38	8	120	78	X
swc1		c.ngle.f	11 1001		39	9	121	79	у
swc2		c.seq.f	11 1010		3a	:	122	7a	Z
		c.ngl.f	11 1011	59	3b	;	123	7b	{
		c.lt.f	11 1100	60	3c	<	124	7c	
sdc1		c.nge.f	11 1101		3d	=	125	7d	}
sdc2		c.le.f	11 1110		3e	>	126	7e	~ DEI
	1 /0/ 1	c.ngt.f	11 1111	63	3f	?	127	7f	DEL
(1) opco	ode(31:26)	== 0							

(2) opcode(31:26) = 17_{ten} (11_{hex}); if fmt(25:21)= 16_{ten} (10_{hex}) f = s (single); if $fmt(25:21) = 17_{ten} (11_{hex}) f = d (double)$

IEEE 754 FLOATING-POINT STANDARD

(3)

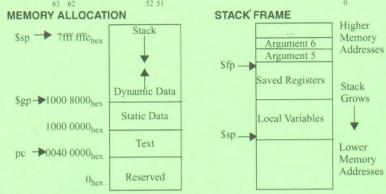
 $(-1)^S \times (1 + Fraction) \times 2^{(Exponent - Bias)}$ where Single Precision Bias = 127, Double Precision Bias = 1023.

IEEE Single Precision and Double Precision Formats:

IEEE 754 Symbols Exponent Fraction Object 0 + 0 0 **#**0 ± Denorm 1 to MAX - 1 anything ± Fl. Pt. Num. MAX 0 ±∞ MAX **≠**0 NaN S.P. MAX = 255, D.P. MAX = 2047

(4)

S Exponent 31 S Fraction Exponent



DATA ALIGNMENT

Double Word										
	Wo	rd		Word						
Halfword		Halfword		Half	fword	Halfword				
Byte	Byte	Byte	Byte	Byte	Byte	Byte	Byte			

EXCEPTION CONTROL REGISTERS: CAUSE AND STATUS

Interrupt Exception D Code Mask Pending M LE Interrupt

BD = Branch Delay, UM = User Mode, EL = Exception Level, IE =Interrupt Enable

EXCEPTION CODES

-/	CEPIN	JIN CC	IDES			
The same of	Number	Name	Cause of Exception	Number	Name	Cause of Exception
	0	Int	Interrupt (hardware)	9	Вр	Breakpoint Exception
	4	AdEL	Address Error Exception (load or instruction fetch)	10	RI	Reserved Instruction Exception
	5	AdES	Address Error Exception (store)	11	CpU	Coprocessor Unimplemented
	6	IBE	Bus Error on Instruction Fetch	12	Ov	Arithmetic Overflow Exception
	7	DBE	Bus Error on Load or Store	13	Tr	Trap
	8	Sys	Syscall Exception	15	FPE	Floating Point Exception

SIZE PREFIXES

	PREFIX	SYMBOL	SIZE	PREFIX	SYMBOL	SIZE	PREFIX	SYMBOL	SIZE	PREFIX	SYMBOL
103	Kilo-	К	210	Kibi-	Ki	1015	Peta-	p	250	Pehi-	P3
106	Mega-	М	220	Mebi-	Mi	1018	Exa-	E	260	Exhi-	Ei
109	Giga-	G	230	Gibi-	Gi	1023	Zetta-	Z	270	Zebi-	Zí
1012	Tera-	Т	240	Tebi-	Ti	1024	Yorta-	Y	280	Yobi-	Yi

- Only instructions for unsigned operations need to be implemented. (addu, addiu, subu, sltiu, sltu, slt, srl)
- However, the immediate fields and offset fields for certain instructions are sign extended to allow negative numbers (addiu, sltiu, beq, bne, lw, sw). So you must implement the assembler to read the fields as signed-extended bits.
- Only loads and stores with 4B word need to be implemented.
- The assembler must support decimal and hexadecimal numbers (0x) for the immediate field, and .data section.
- The register name is always "\$n" n ranges from 0 to 31.
- la (load address) is a pseudo instruction; it should be converted to one or two assembly instructions.

la \$2, VAR1 : VAR1 is a label in the data section

→ It should be converted to lui and ori instructions.

lui \$register, upper 16bit address ori \$register, lower 16bit address

If the lower 16bit address is 0x0000, the ori instruction is useless.

Case1) load address is 0x1000 0000 lui \$2, 0x1000 Case2) load address is 0x1000 0004 lui \$2, 0x1000 ori \$2, \$2, 0x0004

- Directives

.text

- indicates that following items are stored in the user text segment, typically instructions
- It always starts from <u>0x400000</u>

.data

- indicates that following data items are stored in the data segment
- It always starts from <u>0x10000000</u>

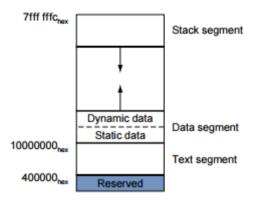
.word

- store n 32-bit quantities in successive memory words

You can assume that the .data and .text directives appear only once, and the .data must appear before .text directive.

Assume that each word in the data section is initialized (Each word has an initial value).

- Memory Layout



- Execution command:

> ./runfile <assembly file>

Your program must produce a single output file (.*o) from the input assembly file (*.s).

- Input format

- Output format

The output of the assembler is an object file which contains a single string of **ASCII** '0' and '1' characters. The ASCII string follows a simplified custom format.

- The first two words (32bits) are the size of text section, and data section.
- The next bytes are the instructions in binary. The length must be equal to the specified text section length.
- After the text section, the rest of bytes are the initial values of the data section.

The following must be the final format of binary ASCII string:

```
<text section size>
<data section size>
<instruction 1>
...
<instruction n>
<value 1>
...
<value m>
```

Please refer to the given examples for a better idea of this format.

2. Program Language

You can choose the programming language among C, and C++. Since subsequent project 2, 3, and 4 should be written in C/C++, you may want to start with C/C++ for the project to get familiar with the language, if you are not yet.

3. GitLab Repository

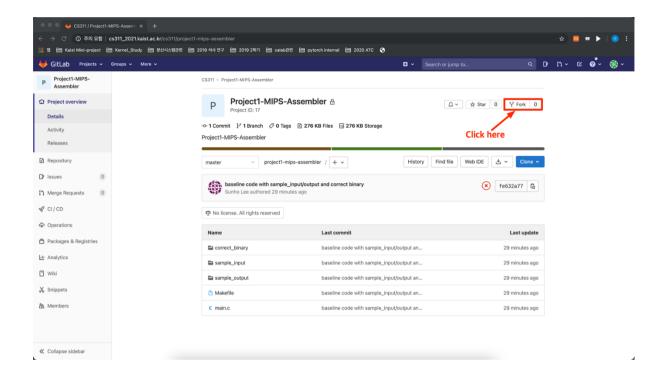
You must create a new repository for this project in GitLab. Please follow the instructions below step-by-step.

First, let's fork the project which the TAs have prepared.

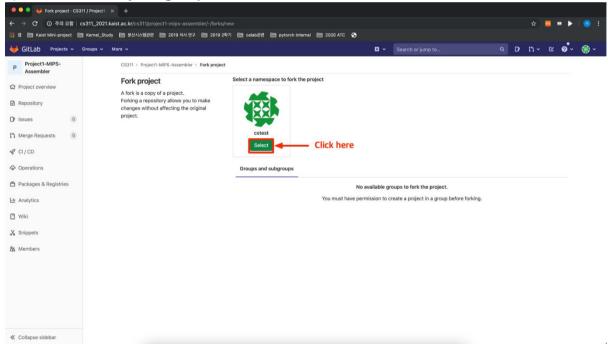
1) Access http://cs311_2021.kaist.ac.kr/cs311/project1-mips-assembler

Examples can be found in the project.

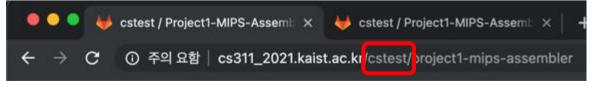
- 2) (If prompted) Login with your GitLab account.
- 3) Click "Fork" (Please refer to the following screen shot).



4) Please choose your group.

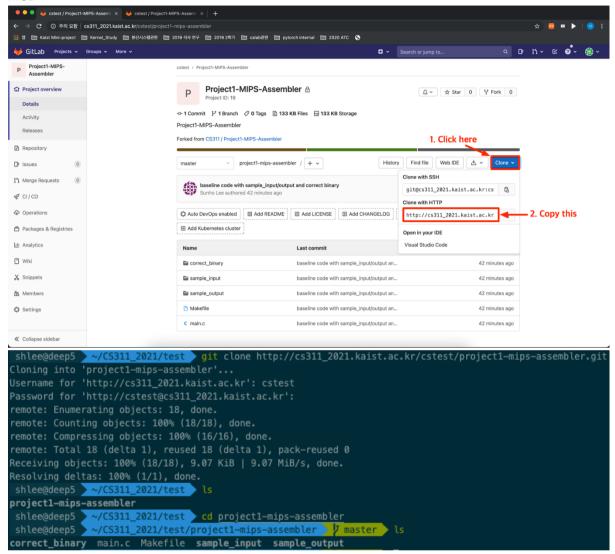


5) Wait for git to import repository and make sure your repository is **made for your group** by checking the URL at the top of your browser. There should be your team's name.



6) Congratulations! You are ready to create a local clone of your repository.

Next, let's make a local copy. In order to work on your project you must make a local copy first.



1) Take a look of the red boxed region and copy it to your clipboard like first figure. Your URL should look similar as the following:

http://cs311_2021.kaist.ac.kr/cstest/project1-mips-assembler.git

- 2) Type the following in your working directory like second figure.
- > git clone YOUR_URL

Example) git clone http://cs311_2021.kaist.ac.kr/cstest/project1-mips-assembler.git

4) If you are successful at making a local copy, you will see a directory called "Project1-MIPS-Assembler". Under this directory you will see the tar file for examples.

Please remember that you are using the local repository as a working space. If you want to commit or submit your work, you must push your work to the remote server. Refer to the link below on how to push your work to the remote repository. (*Focus on 'commit'* and 'push') Link: https://rogerdudler.github.io/git-guide/index.html

4. Grading Policy

Grades will be given based on the 5 examples provided for this project. Your assembler should print the correct corresponding binary code for a given MIPS code.

There are 5 codes to be graded and you will be granted 20% of total score for each correct binary code and **being "Correct" means that every digit and location is the same** to the given output of the example. If a digit is not the same, you will receive **0 score** for that example.

5. Submission (Important!!)

Make sure your code works well on your allocated Linux server.

In fact, it is highly recommended to work on your allocated server throughout this class. Your project will be graded on the same environment as your allocated Linux server.

You must include a Makefile in your submission that builds your 'assembler' into the name 'runfile'. We will be building your assembler using the `make` command. An example Makefile for c is provided in the project directory.

Submit your work to your team's GitLab repository by adding a "submit" tag. Please follow the steps below when submitting.

- 1) Commit and push your code and Makefile to your remote repository.
- 2) Type the following command in your working directory.
- 2-1) > git tag -a submit -m 'whatever message you want'
- 2-2) > git push origin submit

If there is no "submit" tag, your work will not be graded so please remember to submit your work with the tag.

6. Late Policy

You will loss 30% of your score on the first day (April 2^{nd} 0:00 ~ April 2^{nd} 23:59). We will **not** accept works that are submitted after then.

Be aware of plagiarism! Although it is encouraged to discuss with others and refer to extra materials, copying other students or opened code is strictly banned.

The TAs will be comparing your source code with open source codes and other team's code. If you are caught, you will receive a penalty for plagiarism.

If you have any requests or questions regarding administrative issues (such as late submission due to an unfortunate accident, GitLab is not working) please send an e-mail to the TAs(cs311_ta@casys.kaist.ac.kr)