Google techs for developers

JIEUNG KIM

jieungkim@inha.ac.kr

인하대학교
INHA UNIVERSITY

Contents

- Who am I?
- Websites and Youtube channels for Google techs and lives
- Google technologies: Programming language style guides
- Google technologies: Googletest framework
- Google technologies: Machine learning related tools
  - Colab
  - Tensorflow and Keras
  - Tensorflow model optimization toolkit
  - Tensorflow hub and model garden
  - Tensorboard
- Summary

Who am I?

Who am I?

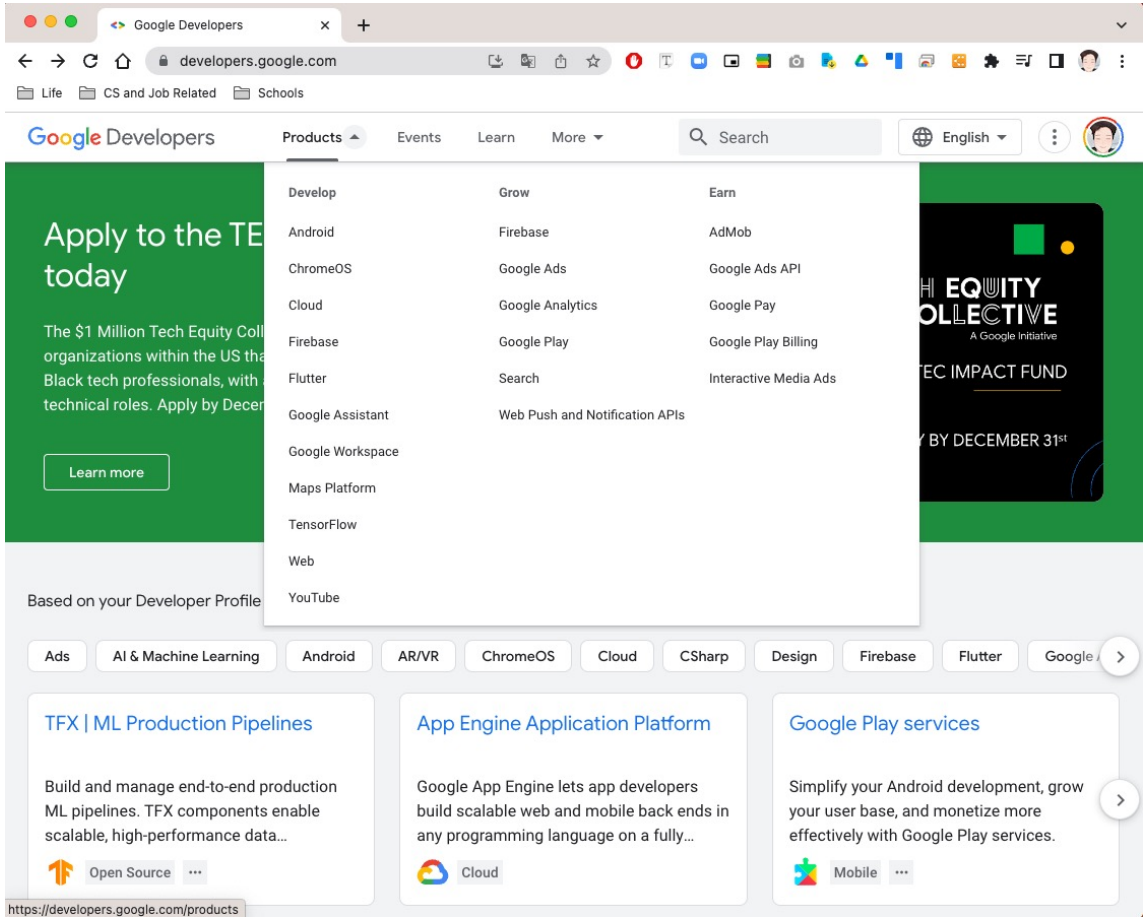
- **Who am I? – https://jieung.kim**
  - I am an **assistant professor** in **Computer Engineering Department**, College of Software and Convergence, Inha University (Incheon, South Korea)
  - Before that, I was a **research engineer** at Google
    - Worked on **machine learning model optimizations** (2022.03~2022.08)
    - Worked on **pKVM formal verification** (2020.05~2022.02)
      - **pKVM**: a new software stack in the Android ecosystem to increase security
      - **Formal verification**: the strongest method to show the correctness and security properties of software with using mathematical and computational logic methods
  - Even before that, I was at Yale University as a **graduate student**, worked on other **formal verification projects**
    - Verify OS & hypervisor based on xv6 (CertiKOS)
    - Provide unified and verified APIs for distributed protocol (ADO)
  - I received my M.S. and B.S. from KAIST and SKKU (South Korea), respectively

Websites and Youtube channels for Google techs and lives

인하대학교
INHA UNIVERSITY

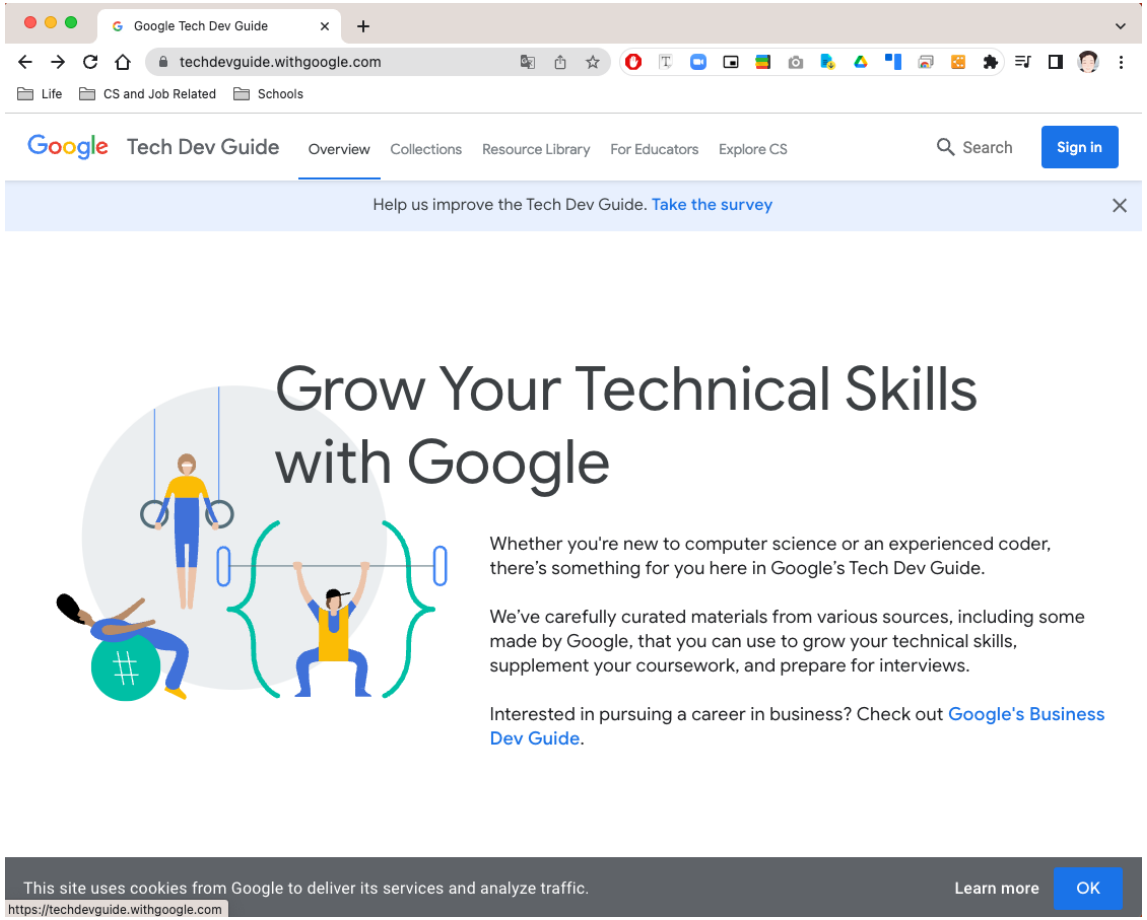
Google developers

https://developers.google.com/



Tech dev guide

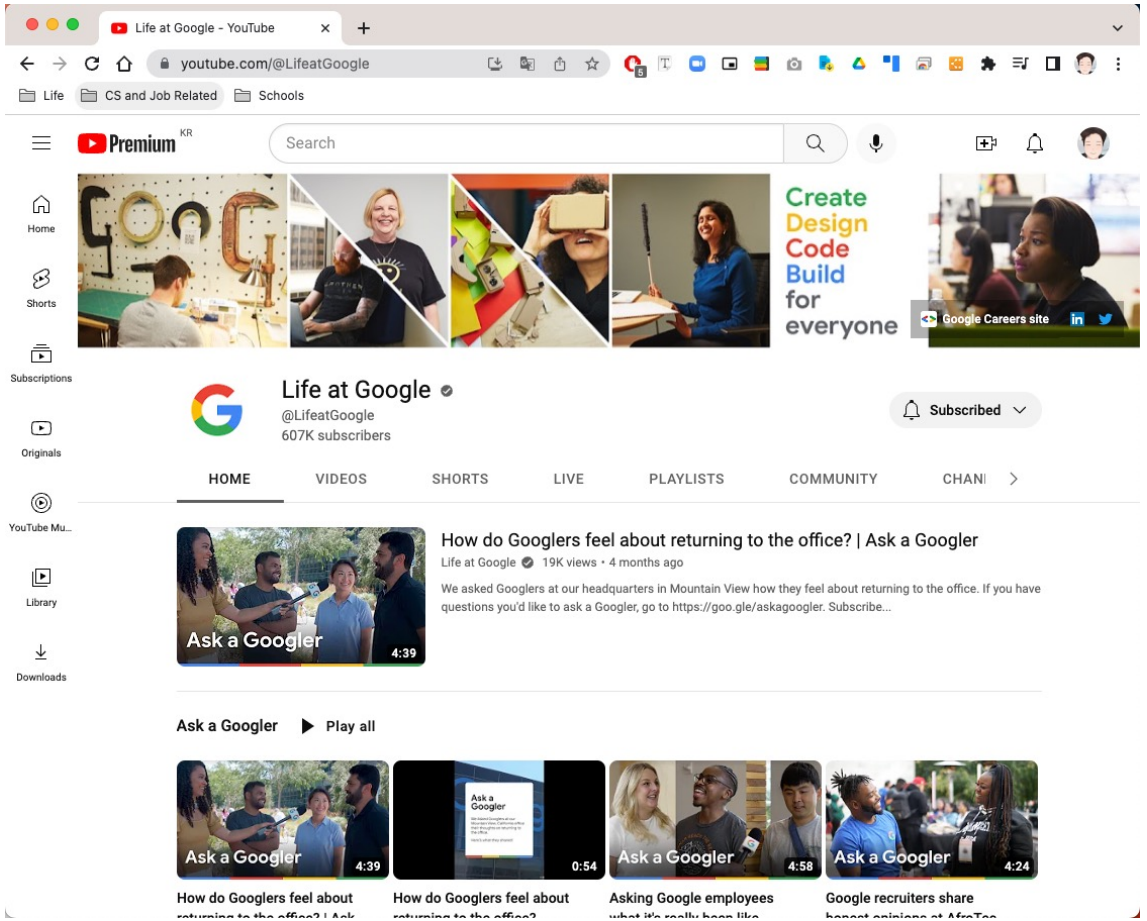
https://techdevguide.withgoogle.com/



인하대학교 INHA UNIVERSITY

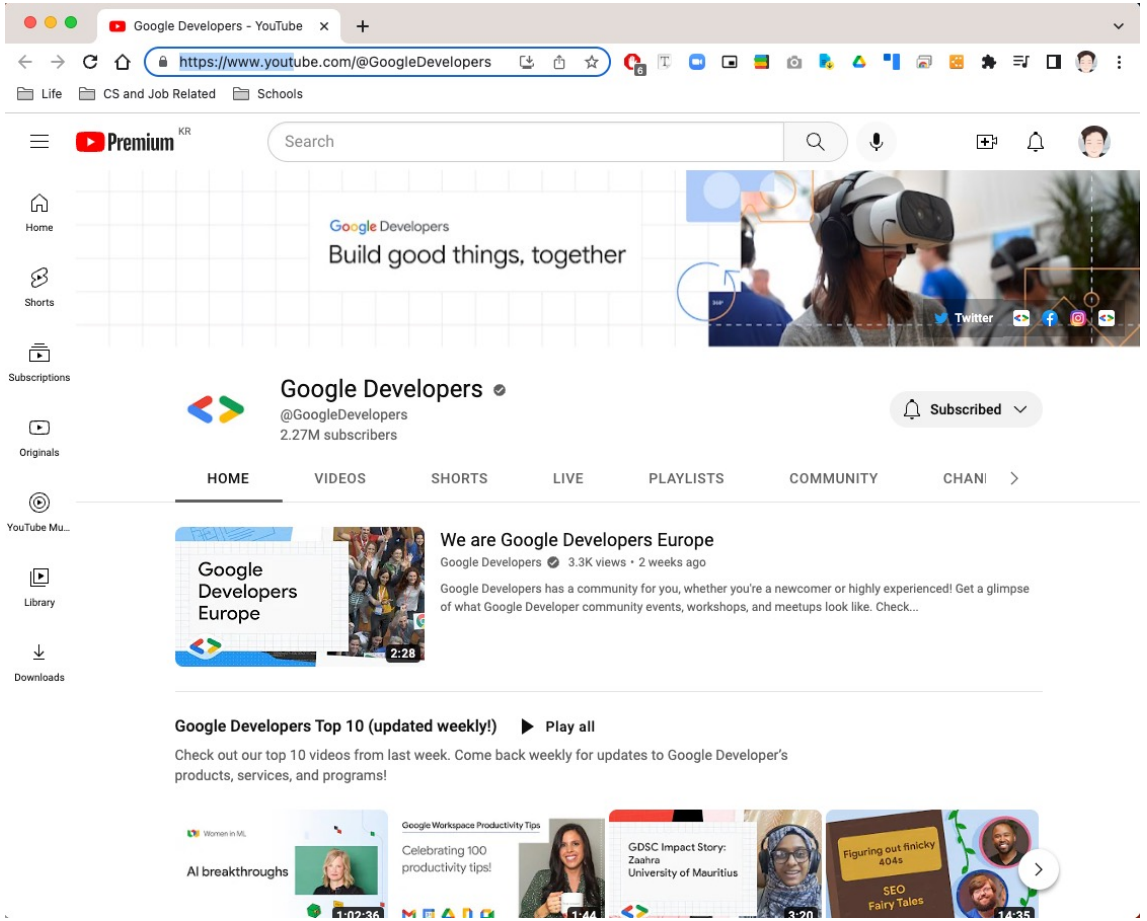
Youtube, Life at Google

https://www.youtube.com/lifeatgoogle



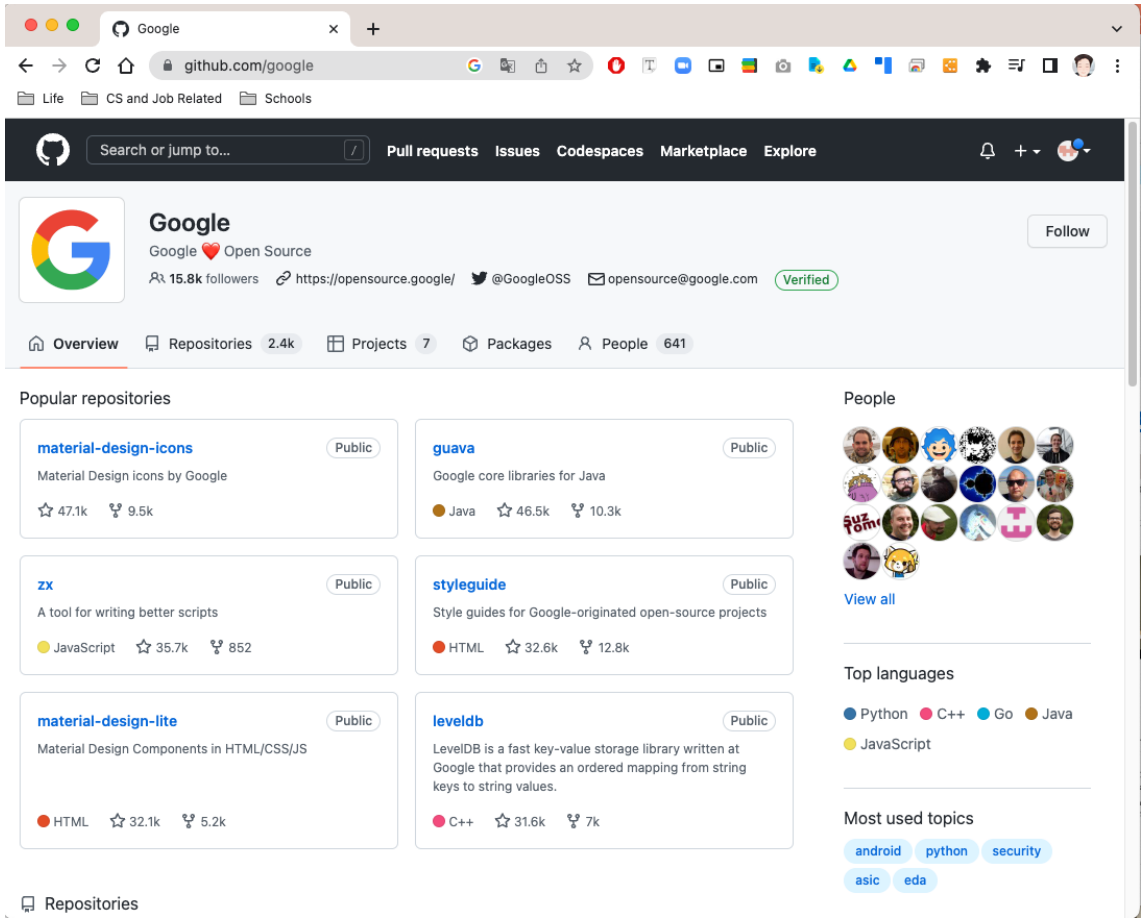
Youtube, Google Developers

https://www.youtube.com/GoogleDevelopers



Google repository at Github

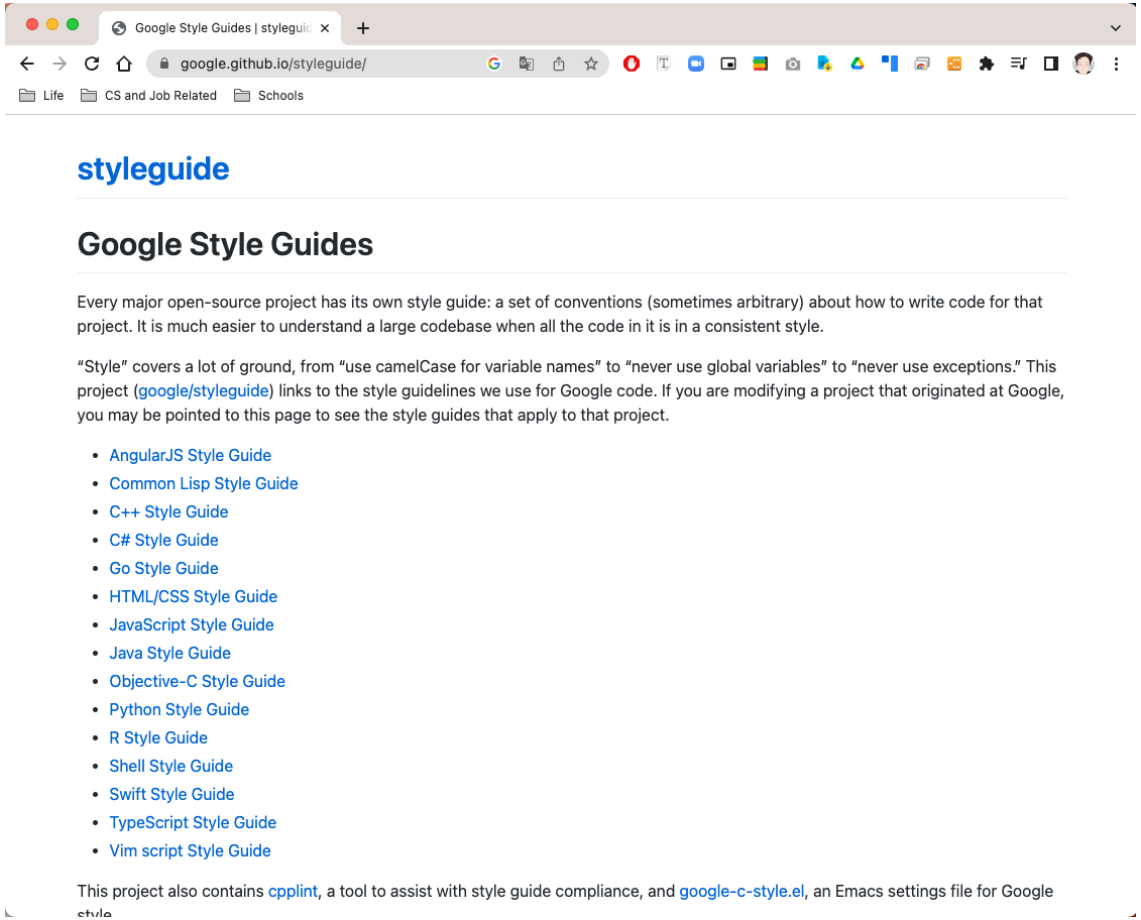
https://github.com/google



인하대학교 INHA UNIVERSITY

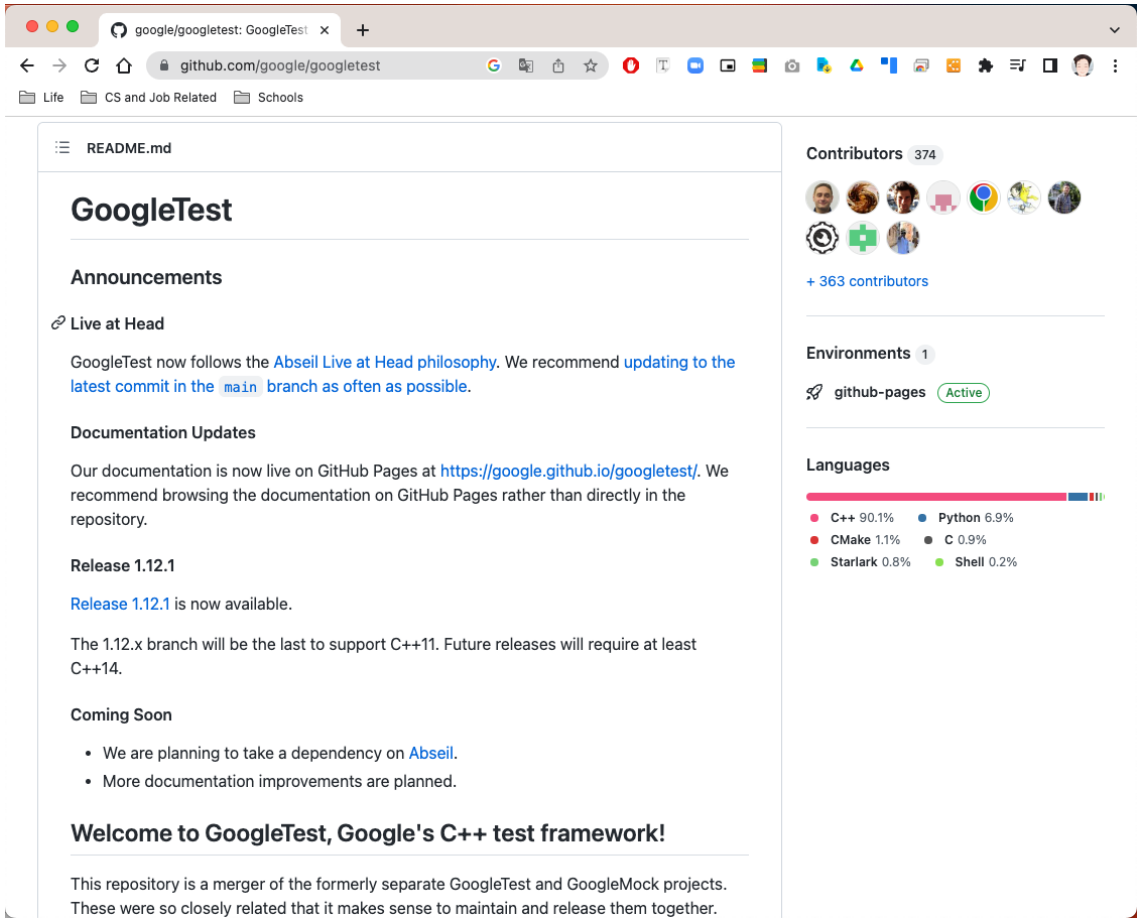
Google style guide

https://google.github.io/styleguide/



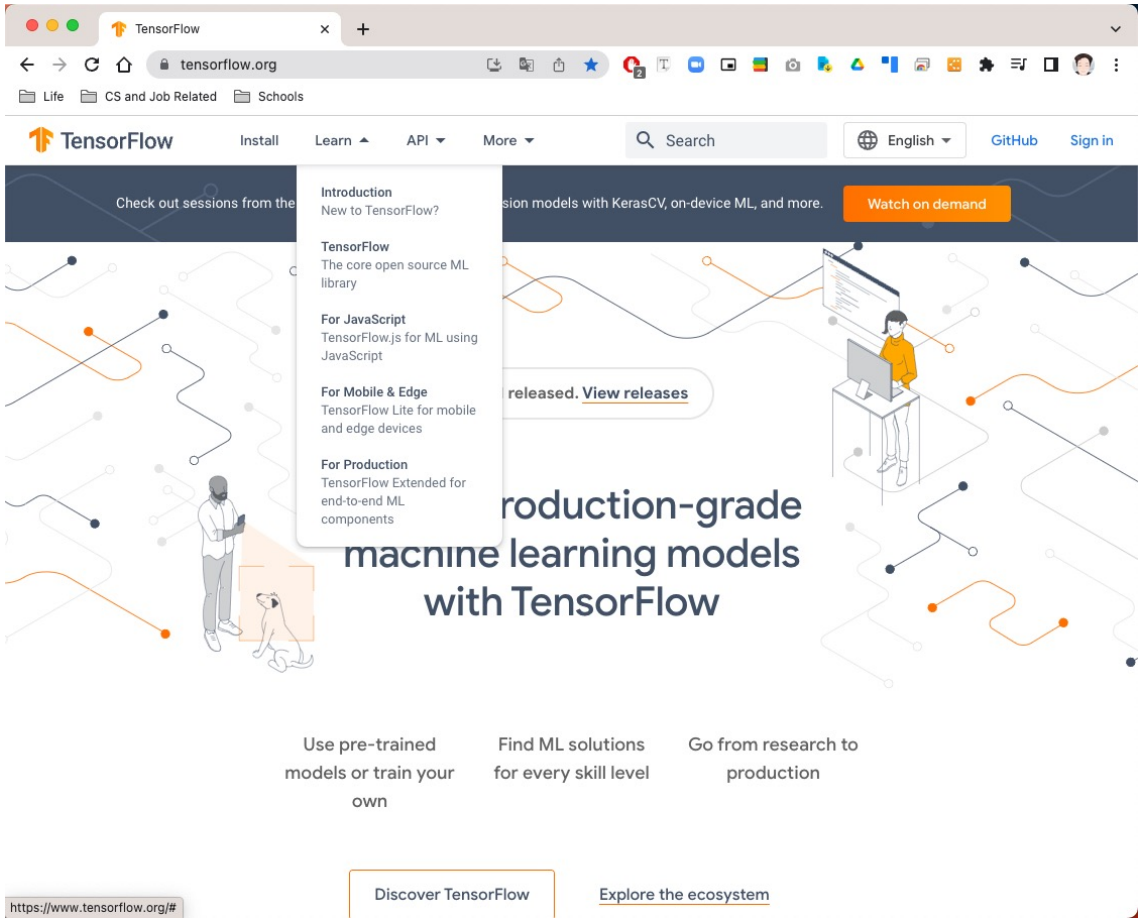
Googletest framework

https://github.com/google/googletest



Tensorflow

https://www.tensorflow.org/



Google technologies: Programming language style guides

Programming language style guides

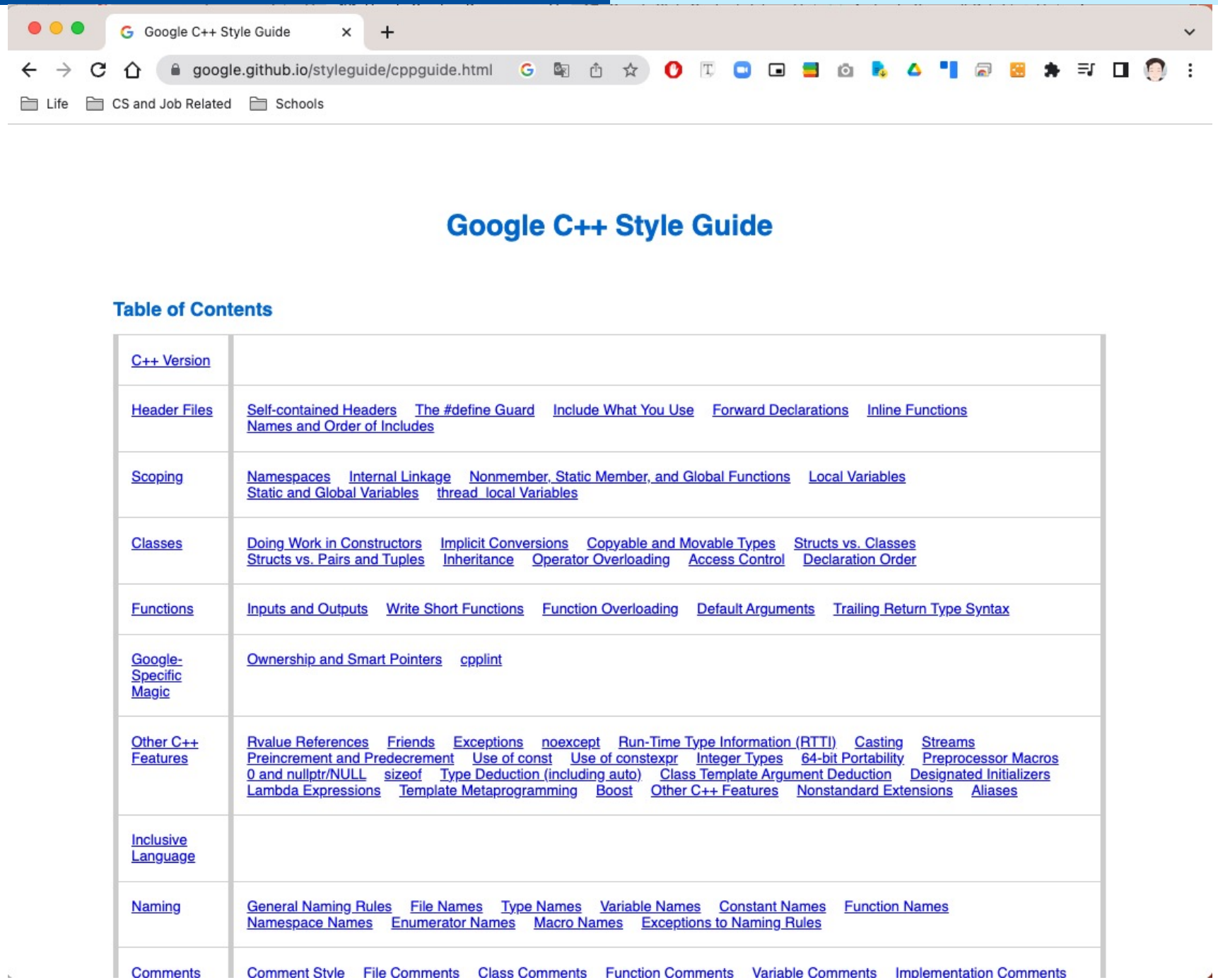
- **Programming language style guides**
  - A guide of programming conventions, style, and best practices for a team or project
  - Following the guides in the development make team's code more consistent and readable
    - Consistent code is easier to read and understand making it faster to add new features
    - **Code review** process is usually check whether the code follows style guides

Code review

- **Code review**
  - Careful, systematic study of source code by people who are not the original author of the code
- **Purpose of code review**
  - Can catch many bugs, design flaws early
  - > 1 person has seen every piece of code
    - Insurance against author's disappearance
  - Forcing function for documentation and code improvements
    - Authors to articulate their decisions
    - Authors participate in the discovery of flaws
    - Prospect of someone reviewing your code raises quality threshold
  - Inexperienced personnel get hands-on experience without hurting code quality
    - Pairing them up with experienced developers
    - Can learn by being a reviewer as well

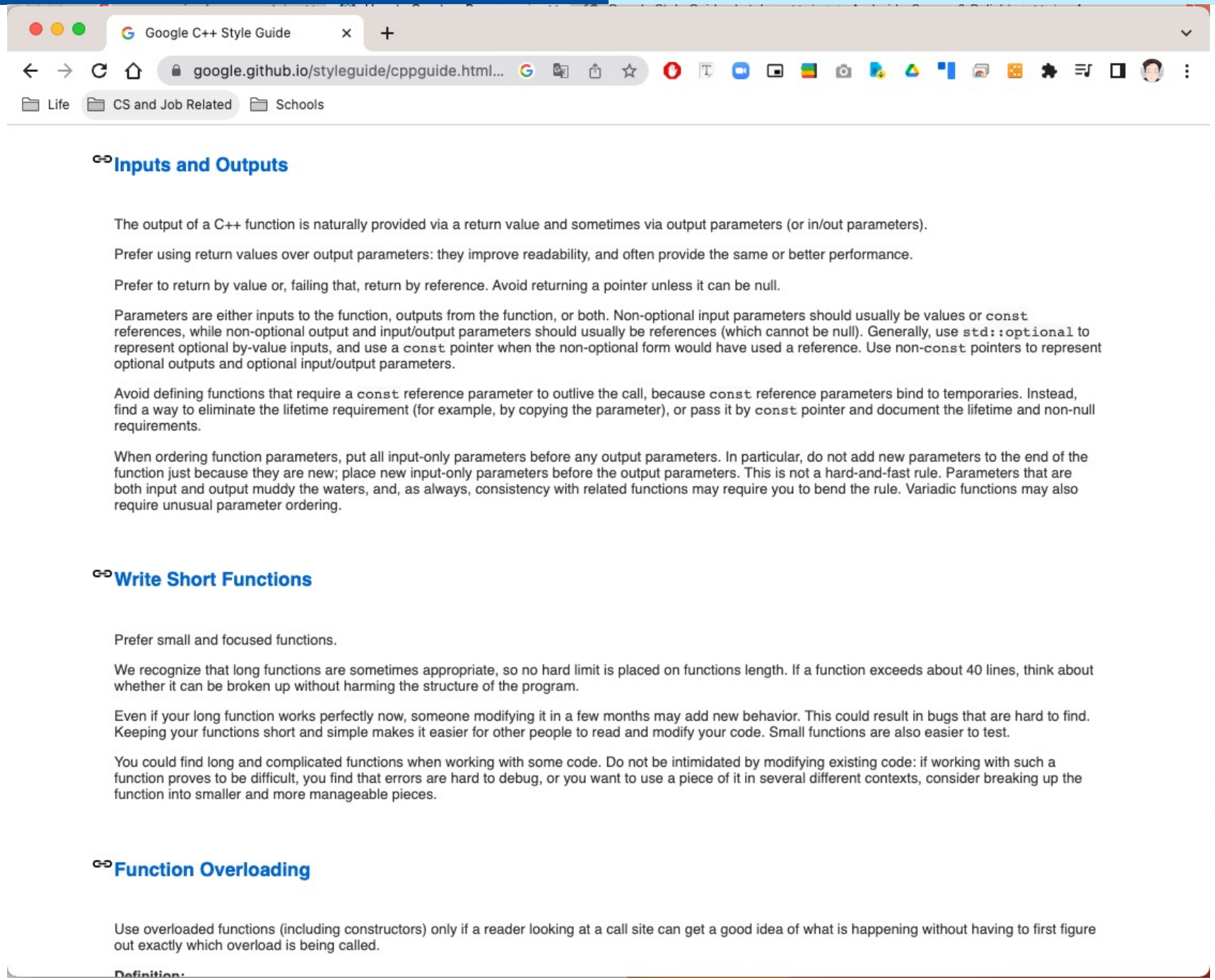
Code review

- **Purpose of code review - by numbers**
  - **From Steve McConnel's Code Complete**
  - Average defect detection rates
    - Unit testing: 25%
    - Function testing: 35%
    - Integration testing: 45%
    - Design and code inspections (reviews): 55% and 60%
  - 11 programs developed by the same group of people
    - First 5 without reviews: average 4.5 errors per 100 lines of code
    - Remaining 6 with reviews: average 0.82 errors per 100 lines of code
    - Errors reduced by > 80%
  - After AT&T introduced reviews, 14% increase in productivity and a 90% decrease in defects



Google C++ Style Guide

Table of Contents



Inputs and Outputs

The output of a C++ function is naturally provided via a return value and sometimes via output parameters (or in/out parameters).

Prefer using return values over output parameters: they improve readability, and often provide the same or better performance.

Prefer to return by value or, failing that, return by reference. Avoid returning a pointer unless it can be null.

Parameters are either inputs to the function, outputs from the function, or both. Non-optional input parameters should usually be values or `const` references, while non-optional output and input/output parameters should usually be references (which cannot be null). Generally, use `std::optional` to represent optional by-value inputs, and use a `const` pointer when the non-optional form would have used a reference. Use non-`const` pointers to represent optional outputs and optional input/output parameters.

Avoid defining functions that require a `const` reference parameter to outlive the call, because `const` reference parameters bind to temporaries. Instead, find a way to eliminate the lifetime requirement (for example, by copying the parameter), or pass it by `const` pointer and document the lifetime and non-null requirements.

When ordering function parameters, put all input-only parameters before any output parameters. In particular, do not add new parameters to the end of the function just because they are new; place new input-only parameters before the output parameters. This is not a hard-and-fast rule. Parameters that are both input and output muddy the waters, and, as always, consistency with related functions may require you to bend the rule. Variadic functions may also require unusual parameter ordering.

Write Short Functions

Prefer small and focused functions.

We recognize that long functions are sometimes appropriate, so no hard limit is placed on functions length. If a function exceeds about 40 lines, think about whether it can be broken up without harming the structure of the program.

Even if your long function works perfectly now, someone modifying it in a few months may add new behavior. This could result in bugs that are hard to find. Keeping your functions short and simple makes it easier for other people to read and modify your code. Small functions are also easier to test.

You could find long and complicated functions when working with some code. Do not be intimidated by modifying existing code: if working with such a function proves to be difficult, you find that errors are hard to debug, or you want to use a piece of it in several different contexts, consider breaking up the function into smaller and more manageable pieces.

Function Overloading

Use overloaded functions (including constructors) only if a reader looking at a call site can get a good idea of what is happening without having to first figure out exactly which overload is being called.

Definition:

인하대학교 INHA UNIVERSITY

Examples

- **Some common guidelines in multiple C++ style guide**
  - The rule of the three: If a class defines one (or more) of the following, it should explicitly define all three, which are 1) destructor, 2) copy constructor, 3) copy assignment operator
  - Do not use #define unless you have to use it
  - Try to use **const** member functions and variables
  - Set up the criteria on class, function, field, and variable names
  - Locate functions in proper classes
  - Try to use initializer list
  - Use iteration over STL containers
  - …

```
#define Fresh 1
#define Sophomore 2
#define Junior 3
#define Senior 4

class Student {
  public:
  Student(int id, int year) {
    student_id = id;
    student_year = year;
  };

  ~Student();

  int GetStudentID() { return student_id; }
  int get_student_year() { return student_year; }

  private:
  int student_id;
  int student_year;
};


bool FindStudent(int id, std::vector<Student> students) {
  for (int i = 0; i < students.size(); i++) {
    if (students[i].GetStudentID() == id) {
      return true;
    }
  }
  return false;
}
```

Do not use #define

Initializer list is not used

Violate the rule of the three

Inconsistency in function names

Fields are not distinguishable from local variables

References should be used

Iterator is not used

```
class Student {
  public:
  enum StudentYear { FRESH = 1, Sophomore, Junior, Senior };

  Student(const int id, const StudentYear year) :
    id_(id), year_(year) {};
  Student(const Student& student) :
    id_(student.id_), year_(student.year_) {};
  Student& operator=(const Student& student) {
    if (this != &student) {
      *this = Student(student);
    }
    return *this;
  };
  ~Student();

  int GetId() { return id_; }
  int GetYear() { return year_; }

  bool FindStudent(const int id,
    const std::vector<Student>& students) const {
    for (const auto& student : students) {
      if (student.id_ == id) {
        return true;
      }
    }
    return false;
  }

  private:
  const int id_;
  StudentYear year_;
};
```

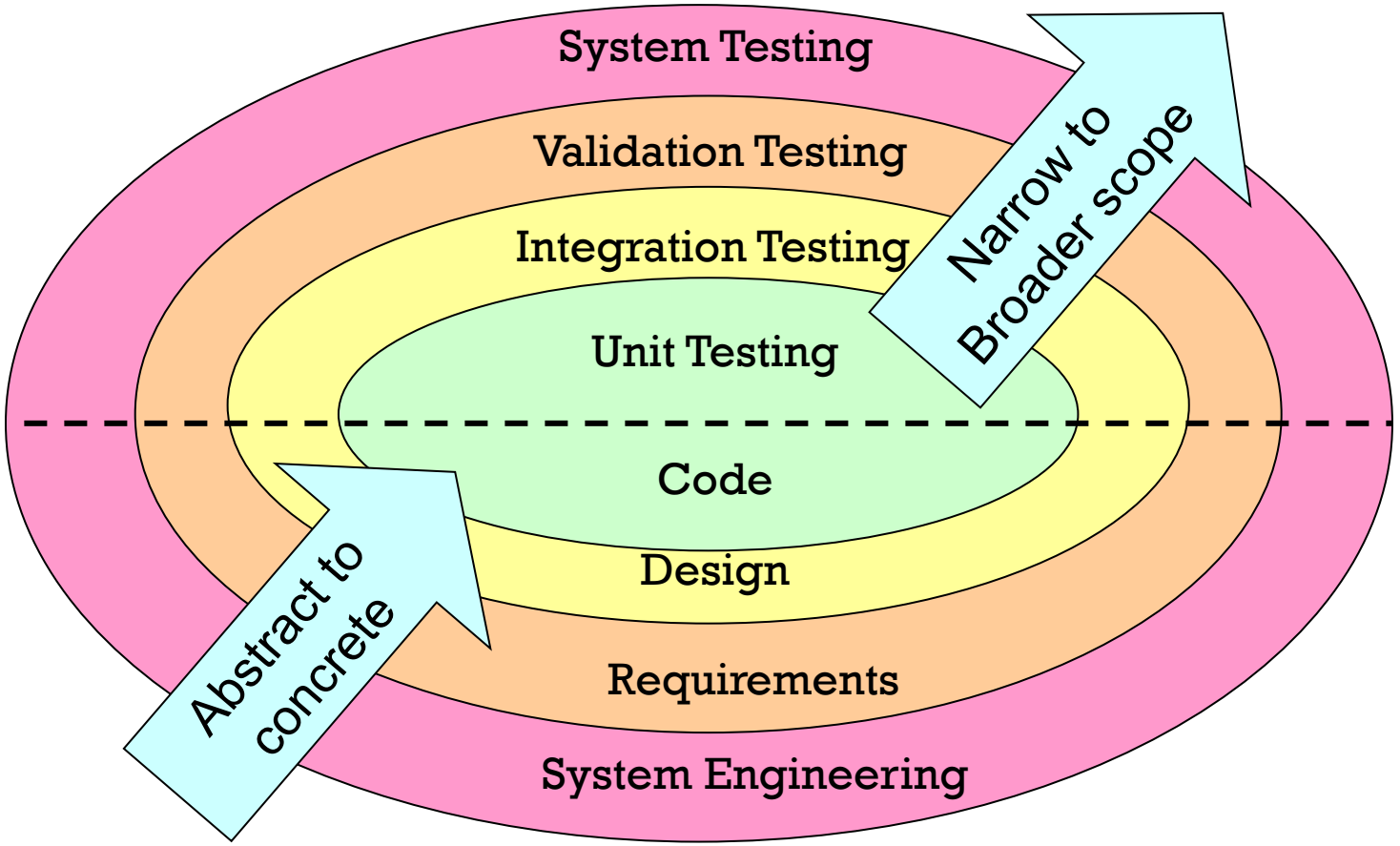
Google technologies: GoogleTest Framework

Software testing

- **Software testing**
  - Evaluation of the software against requirements gathered from users and system specifications
- **Does testing really work?**
  - "measuring over 20 projects: if you have a large number of unit tests your code will be **an order of magnitude (x10)** less complex.“
  - **Controlled study results**:
    - “*..quality increased linearly with the number of programmer tests...*”
    - “*..test-first students on average wrote more tests and, in turn, students who wrote more tests tended to be more productive...*”

http://agilepainrelief.com/notesfromatooluser/2008/11/misconceptions-with-test-driven-development.html
http://collaboration.csc.ncsu.edu/laurie/Papers/TDDpaperv8.pdf

Software testing



Software testing

- **Unit testing frameworks and libraries**
  - **Java**
    - NUnit, **Junit**, TestNG, Mockito, and PHPUnit
  - **Python**
    - Robot, **PyTest**, **Unittest**, DocTest, Nose2, and Testify
  - **C/C++**
    - **Googletest**, Boot Test Library, QA Systems Cantata, Parasoft C/C++ test, Microsoft Visual Studio, Cppunit, Catch, Bandit, and CppUTest
  - **JavaScript**
    - Jest, Mocah, Storybook, Jasmine, Cypress, Puppeteer, Testing Library, and WebdriverIO

Googletest framework

- **Googletest framework**
  - A **unit testing library** for the C++ programming language.
  - **Repository**
    - http://code.google.com/p/googletest/
  - **Projects using Google Test**
    - Android open source project operating system
    - Chromium projects (behind the Chrome browser, Edge browser, and Chrome OS)
    - LLVM compiler
    - Protocol Buffers (Google's data interchange format)
    - OpenCV computer vision library
    - Several internal C++ projects at Google
  - **Study materials**
    - README file: https://github.com/google/googletest/blob/master/README.md
    - Googletest user's guide: https://google.github.io/googletest/
    - Whittaker, James (2012). How Google Tests Software. Boston, Massachusetts: Pearson Education. ISBN 0-321-80302-7

Create tests

- **Creating a basic test**
  - **Target code: prototype for square-root**

```
double square-root (const double);
```

  - **Test case with Googletest**

```
#include "gtest/gtest.h"
TEST (SquareRootTest, PositiveNos) {
    EXPECT_EQ (18.0, square-root (324.0));
    EXPECT_EQ (25.4, square-root (645.16));
    EXPECT_EQ (50.3321, square-root (2533.310224));
}
TEST (SquareRootTest, ZeroAndNegativeNos) {
    ASSERT_EQ (0.0, square-root (0.0));
    ASSERT_EQ (-1, square-root (-22.0));
}
```

Create tests

Predefined macro in `gtest.h`

Test hierarchy name

Unit test name

```
double square-root (const double);
```

```
#include "gtest/gtest.h"
TEST (SquareRootTest, PositiveNos) {
    EXPECT_EQ (18.0, square-root (324.0));
    EXPECT_EQ (25.4, square-root (645.16));
    EXPECT_EQ (50.3321, square-root (2533.310224));
}
TEST (SquareRootTest, ZeroAndNegativeNos) {
    ASSERT_EQ (0.0, square-root (0.0));
    ASSERT_EQ (-1, square-root (-22.0));
}
```

Predefined macros that checks result of `square-root`

Check results

- Basic assertions

Fatal assertion	Nonfatal assertion	Verifies
`ASSERT_TRUE(condtion);`	`EXPECT_TRUE(condtion);`	`Condition` **is true**
`ASSERT_FALSE(condition);`	`EXPECT_FALSE(condition);`	`Condition` **is false**

Check results

- Binary comparison

Fatal assertion	Nonfatal assertion	Verifies
`ASSERT_EQ(expected, actual);`	`EXPECTED_EQ(expected, actual);`	`expected == actual`
`ASSERT_NE(val1, val2);`	`EXPECT_NE(val1, val2);`	`val1 != val2`
`ASSERT_LT(val1, val2);`	`EXPECT_LT(val1, val2);`	`val1 < val2`
`ASSERT_LE(val1, val2);`	`EXPECT_LE(val1, val2);`	`val1 <= val2`
`ASSERT_GT(val1, val2);`	`EXPECT_GT(val1, val2);`	`val1 > val2`
`ASSERT_GE(val1, val2);`	`EXPECT_GE(val1, val2);`	`val1 >= val2`

Run tests

- Initialize the framework
- Must be called before `RUN_ALL_TESTS`

```
int main(int argc, char **argv) {
    ::testing::InitGoogleTest(&argc, argv);
    return RUN_ALL_TESTS();
}
```

- Must be called only once
  - Multiple calls to it conflicts some features of the framework
- Automatically detects and runs all test tests defined using the `TEST` macro

Run tests

```
Running main() from user_main.cpp
[==========] Running 2 tests from 1 test case.
[----------] Global test environment set-up.
[----------] 2 tests from SquareRootTest
[ RUN      ] SquareRootTest.PositiveNos
..\user_sqrt.cpp(6862): error: Value of: sqrt (2533.310224)
  Actual: 50.332
Expected: 50.3321
[  FAILED  ] SquareRootTest.PositiveNos (9 ms)
[ RUN      ] SquareRootTest.ZeroAndNegativeNos
[       OK ] SquareRootTest.ZeroAndNegativeNos (0 ms)
[----------] 2 tests from SquareRootTest (0 ms total)

[----------] Global test environment tear-down
[==========] 2 tests from 1 test case ran. (10 ms total)
[  PASSED  ] 1 test.
[  FAILED  ] 1 test, listed below:
[  FAILED  ] SquareRootTest.PositiveNos

1 FAILED TEST
```

Google technologies:
Machine learning related tools

Machine learning related tools

- **Machine learning related tools**
  - Colab: Allows you to write and execute Python in your browser, with zero configuration required, access to GPUs free of charge, and easy sharing
  - Tensorflow and Keras
    - Tensorflow : A free and open-source software library for machine learning and artificial intelligence
    - Keras (an interface for the Tensorflow library): An open-source software library that provides a Python interface for artificial neural networks
  - Tensorflow hub and model garden: Pre-built tensorflow models
  - Tensorflow model optimization toolkit: A suite of tools for optimizing ML models for deployment and execution
  - Tensorboard: Provides the visualization and tooling needed for machine learning experimentation

Colab

- **Colab**
  - A project from Google Research, a free, Jupyter based environment
  - It allows us to create Jupyter programming notebooks to write and execute Python in a web browser
    - It also supports other Python-based third-party tools and machine learning frameworks such as Pandas, PyTorch, Tensorflow, Keras, Monk, OpenCV, and others
  - Google provides the use of free GPU for your Colab notebooks

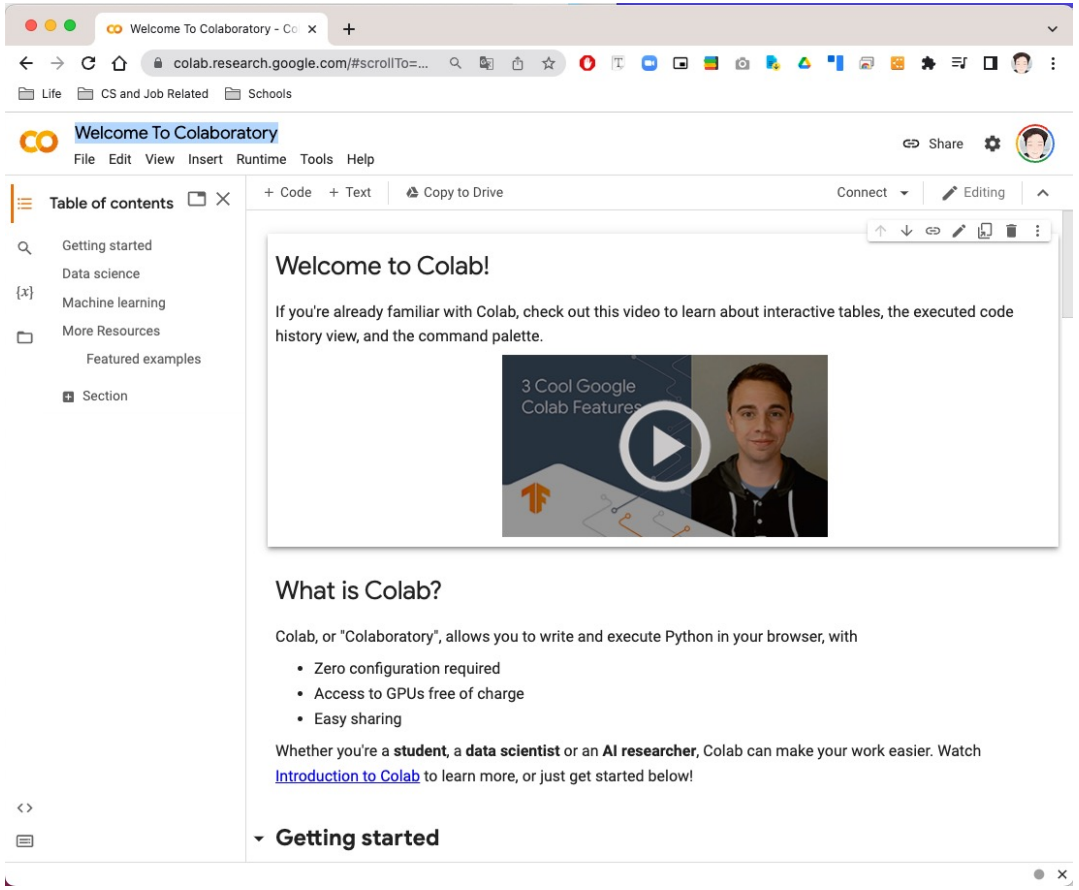
Colab

- **Introduction to Colab**

인하대학교
INHA UNIVERSITY

Colab

- **Welcome To Colab (https://colab.research.google.com/)**

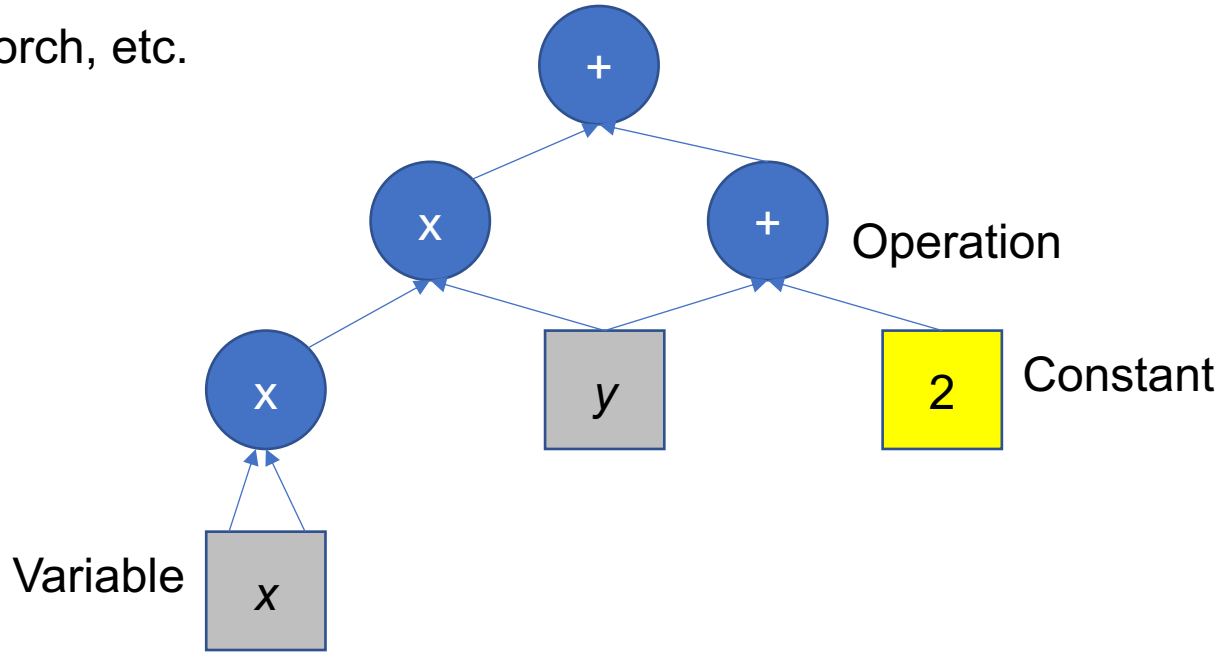


Tensorflow and Keras

- **Deep learning frameworks**

Tensorflow and Keras

- **Deep learning frameworks**
  - Tools for defining static or dynamic general-purpose computational graphs
  - Seamless CPU / GPU usage
  - Python / numpy or R interfaces instead of C, C++, CUDA or HIP
  - Open source
  - Tensorflow, PyTorch, etc.

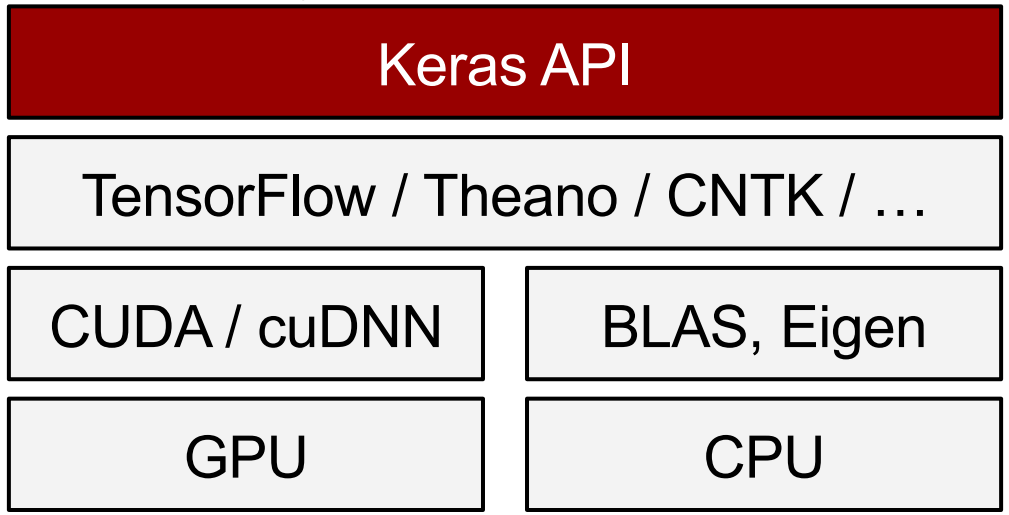


Tensorflow and Keras

- **Tensorflow (https://www.tensorflow.org/)**
  - Deep learning frameworks from Google (The initial version was the name with DistBelief)
  - It can be used in a wide variety of programming languages (e.g., Python, JavaScript, C++, and Java)
  - Companies using Tensorflow
    - Google : Translate, Google Brain's Magenta
    - DeepMind : WaveNet Text to Speech

Tensorflow and Keras

- **Keras**
  - A high-level neural networks API
  - A frontend API for Tensorflow 2.0
    - https://keras.io/
    - https://www.tensorflow.org/guide/keras
  - It dramatically increase the usability of Tensorflow



Tensorflow and Keras

- **Tensorflow 1.0 and Tensorflow 2.0**

```
#Tensorflow xor
import tensorflow as tf
import numpy as np

learning_rate = 0.1
X_train = [[0, 0], [0, 1], [1, 0], [1, 1]]

Y_train = [[0], [1], [1], [0]]

X_train = np.array(X_train, dtype=np.float32)
Y_train = np.array(Y_train, dtype=np.float32)

X = tf.placeholder(tf.float32, [None, 2])
Y = tf.placeholder(tf.float32, [None, 1])

W = tf.Variable(tf.random_normal([2, 1]), name='weight')
b = tf.Variable(tf.random_normal([1]), name='bias')

hypothesis = tf.sigmoid(tf.matmul(X, W) + b)
cost = -tf.reduce_mean(Y * tf.log(hypothesis) + (1 - Y) * tf.log(1 - hypothesis))
train = tf.train.GradientDescentOptimizer(learning_rate=learning_rate).minimize(cost)
predicted = tf.cast(hypothesis > 0.5, dtype=tf.float32)
accuracy = tf.reduce_mean(tf.cast(tf.equal(predicted, Y), dtype=tf.float32))

with tf.Session() as sess:
    sess.run(tf.global_variables_initializer())
    for step in range(10001):
        sess.run(train, feed_dict={X: X_train, Y: Y_train})
        if step % 100 == 0:
            print(step, sess.run(cost, feed_dict=
            {X: X_train, Y: Y_train}), sess.run(W))
    h, c, a = sess.run([hypothesis, predicted, accuracy], feed_dict={X: X_train, Y: Y_train})
    print("\nHypothesis: ", h, "\nCorrect: ", c, "\nAccuracy: ", a)
```

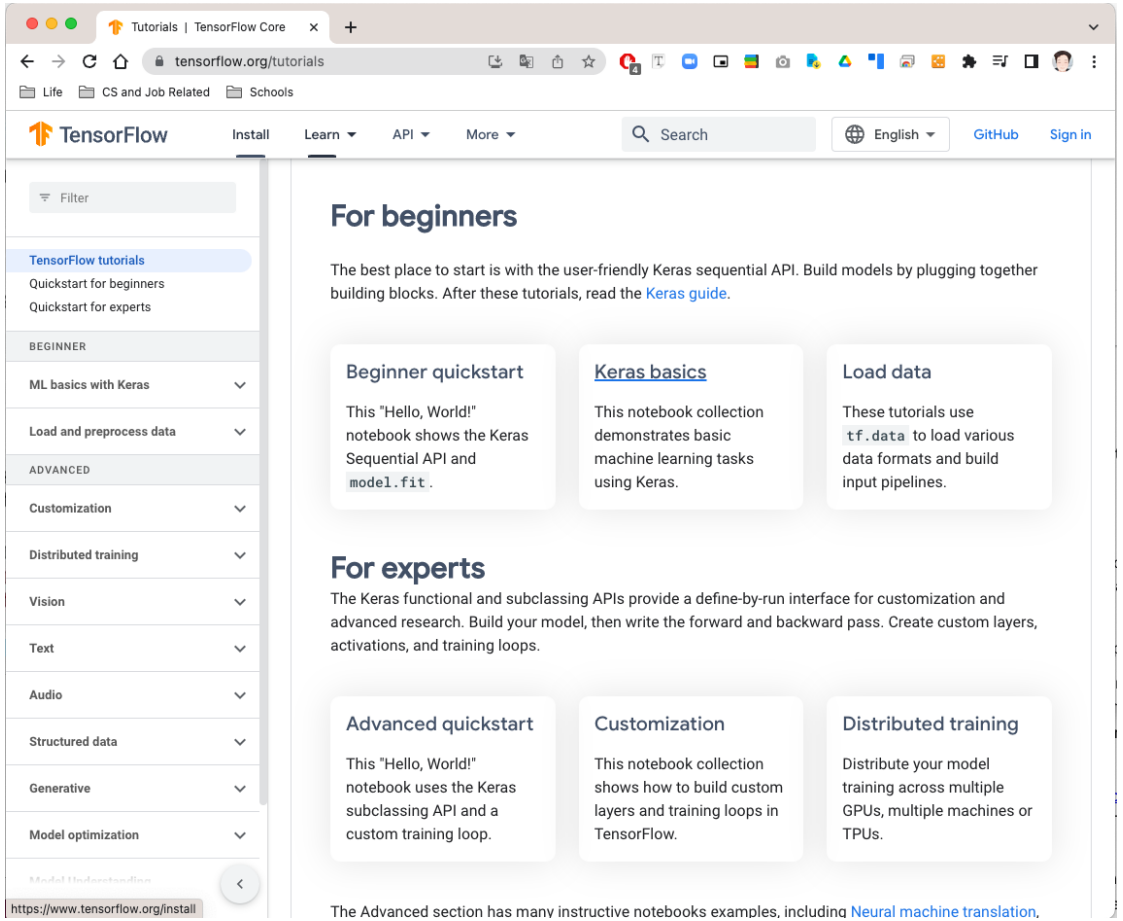
Tensorflow and Keras

- **Tensorflow 1.0 and Tensorflow 2.0**

```
#Tensorflow2 Keras Xor
import numpy as np
import tensorflow as tf

X_train = np.array([[0, 0], [0, 1], [1, 0], [1, 1]], 'float32')
Y_train = np.array([[0], [1], [1], [0]], 'float32')

model=tf.keras.Sequential()
model.add(tf.keras.layers.Dense(4, imput_dim=2))
model.add(tf.keras.layers.Activation('sigmoid'))
model.add(tf.keras.layers.Dense(1))
model.add(tf.keras.layers.Activation('sigmoid'))

sgd=tf.keras.optimizers.SGD(lr=0.1)
model.compile(loss='binary_crossentropy', optimizer=sgd)
model.fit(X_train, Y_train, batch_size=1, epochs=2000)

_predict = model.predict_proba(X_train)
print('predict', _predict)
print('result=', np.array(np.array(_predict) > 0.5, np.int))
```

Tensorflow and Keras

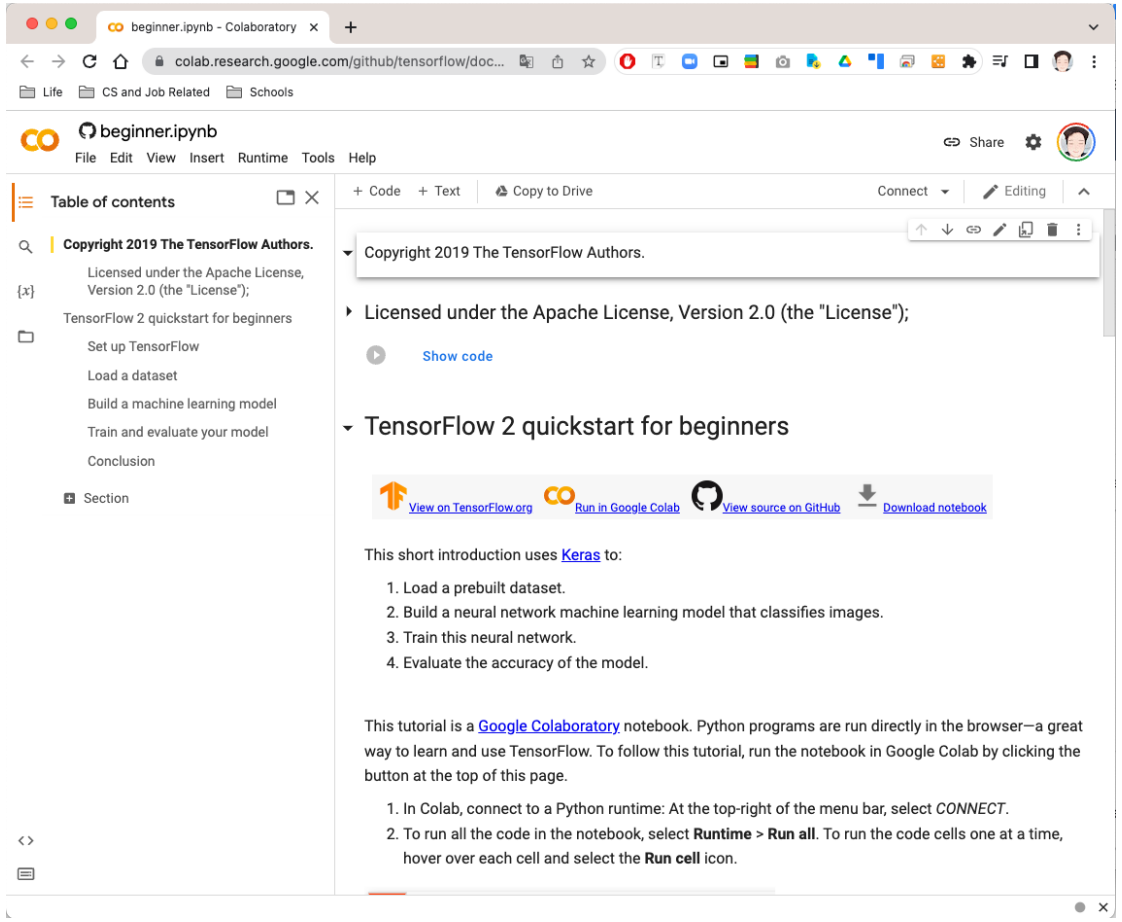
https://www.tensorflow.org/tutorials



인하대학교 INHA UNIVERSITY

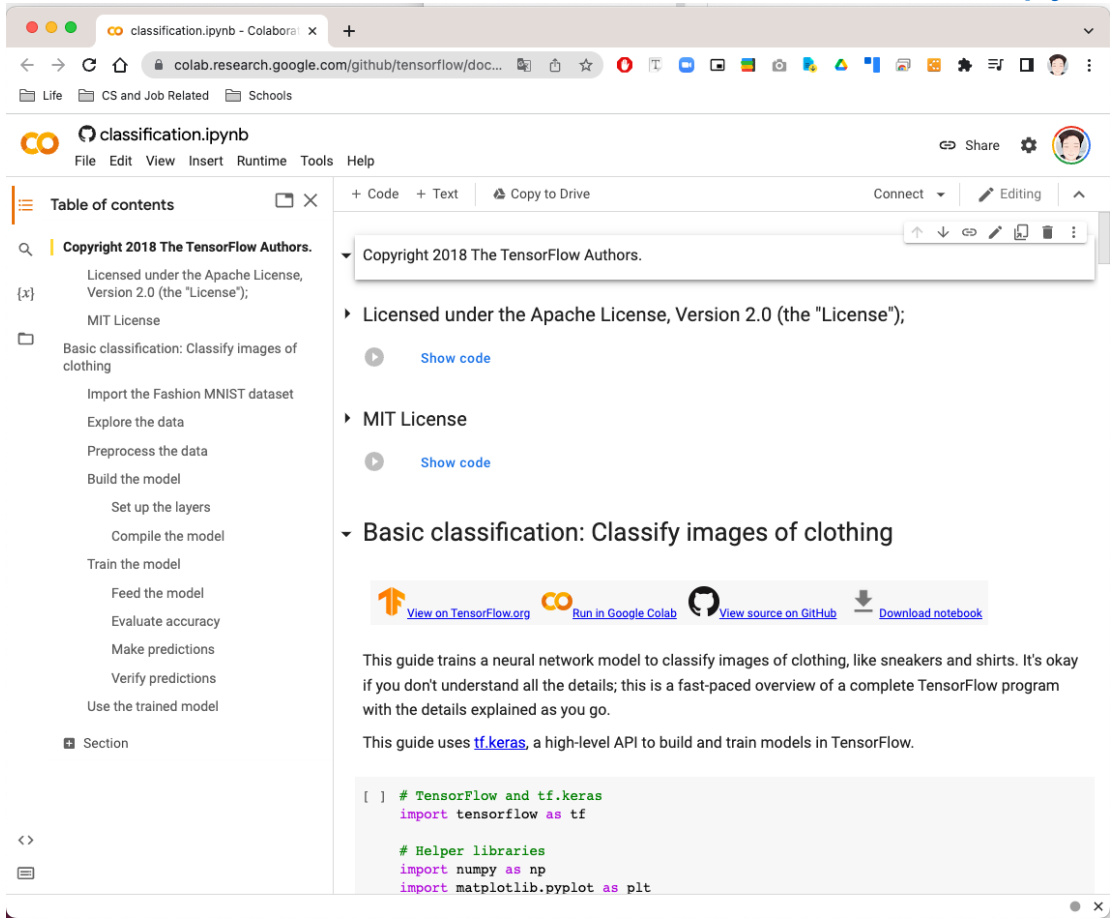
Tensorflow and Keras

https://www.tensorflow.org/tutorials/quickstart/beginner



Tensorflow and Keras

https://colab.research.google.com/github/tensorflow/docs/blob/master/site/en/tutorials/keras/classification.ipynb



Tensorflow and Keras

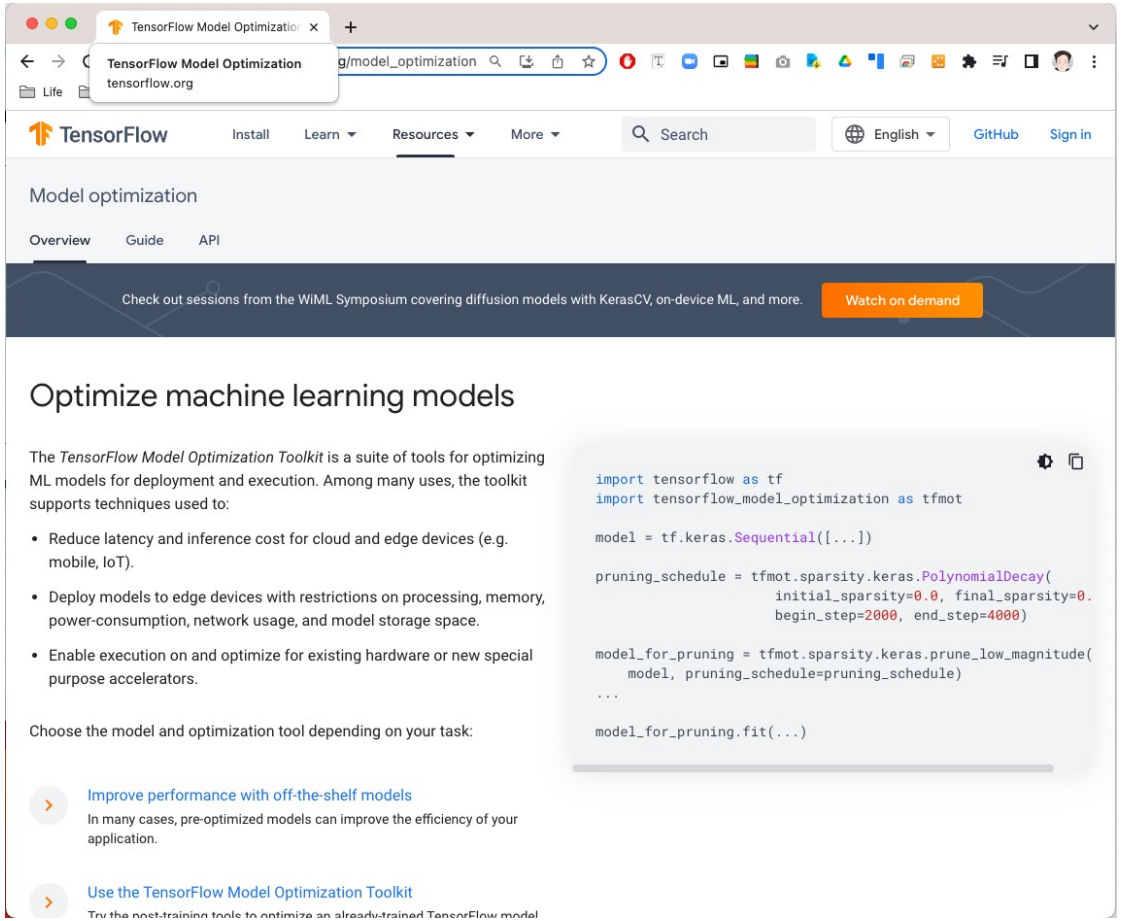
- **Tensorflow hub and model garden**
  - Tensorflow Hub (https://www.tensorflow.org/hub)
    - A repository of trained machine learning models ready for fine-tuning and deployable anywhere
    - Reuse trained models like BERT and Faster R-CNN with just a few lines of code
  - Model garden (https://www.tensorflow.org/guide/model_garden)
    - Provides implementations of many state-of-the-art machine learning (ML) models for vision and natural language processing (NLP)
    - Provides workflow tools to let you **quickly configure** and run those models on **standard datasets**

Tensorflow model optimization toolkit

- **Optimize machine learning models**
  - A suite of tools for optimizing ML models for deployment and execution
  - The toolkit supports techniques used to:
    - Reduce latency and inference cost for cloud and edge devices (e.g., mobile, IoT)
    - Deploy models to edge devices with restrictions on processing, memory, power-consumption, network usage, and model storage space
    - Enable execution on and optimize for existing hardware or new special purpose accelerators
  - Two famous techniques
    - Weight pruning
    - **Quantization**
    - Weight clustering

Tensorflow model optimization toolkit

https://www.tensorflow.org/model_optimization



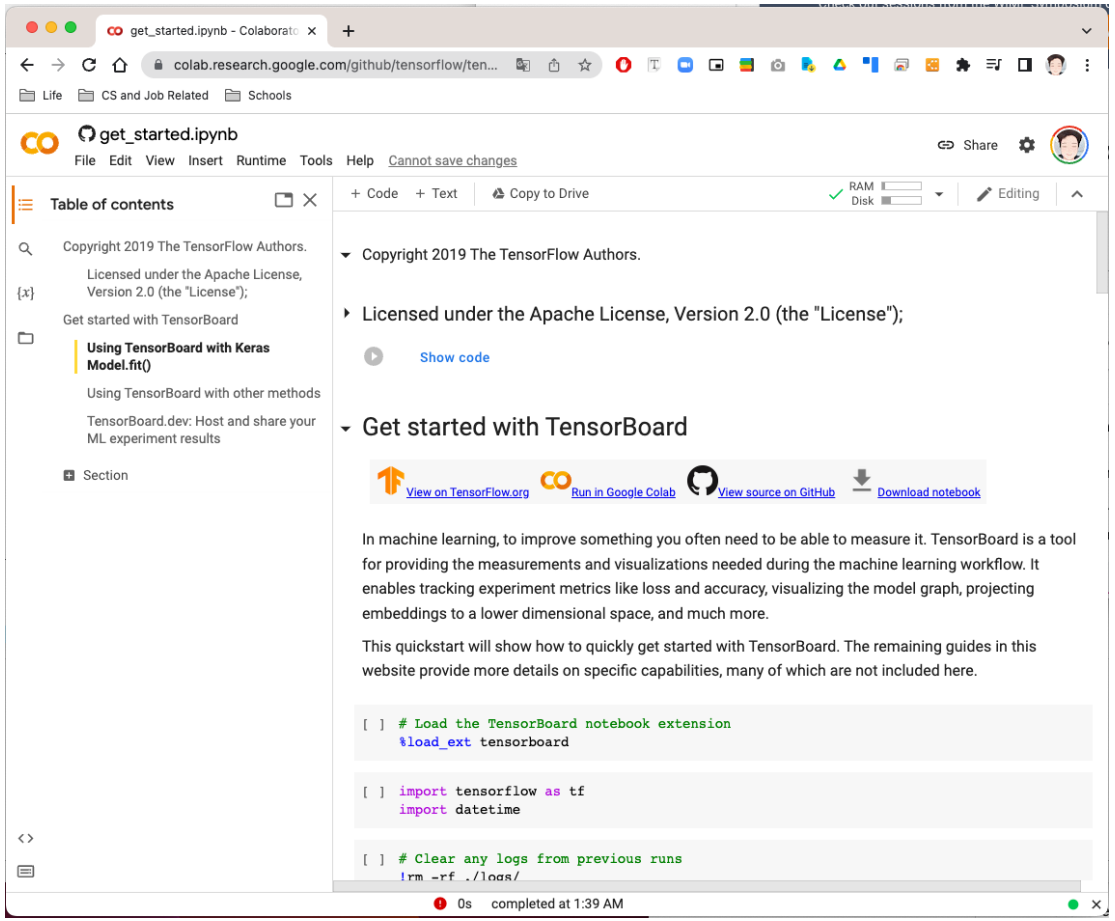
인하대학교 INHA UNIVERSITY

Tensorboard

- **Tensorboard (https://www.tensorflow.org/tensorboard)**
  - provides the visualization and tooling needed for machine learning experimentation:
    - Tracking and visualizing metrics such as loss and accuracy
    - Visualizing the model graph (ops and layers)
    - Viewing histograms of weights, biases, or other tensors as they change over time
    - Projecting embeddings to a lower dimensional space
    - Displaying images, text, and audio data
    - Profiling Tensorflow programs
    - And much more

Tensorboard

https://www.tensorflow.org/tensorboard/get_started



Summary

Summary

- **Summary**
  - Websites and Youtube channels for Google techs and lives
    - Provides a lot of useful information for Google and google techs
  - Programming language style guides and Googletest framework
    - Following the style guide and use a Googletest framework (only for C++) will dramatically improve the project quality
  - Machine learning related tools
    - Colab: A web-based and highly readable Python programming IDE
    - Tensorflow and Keras: A tool to build machine learning models
    - Tensorflow model optimization toolkit: A tool to optimize machine learning models
    - Tensorflow hub and model garden: Pre-built machine learning models
    - Tensorboard: A tool to experiment machine learning models

Q & A