

# 2017 清華大學產業長期實習計畫



## Logitech 實習報名表

姓 名	林杰	英文姓名	Chieh Lin(Hubert)
系 級	資工大三(第四年)	學號	102000039
性 別	<input checked="" type="checkbox"/> 男 <input type="checkbox"/> 女	出生日期	民國 84 年 05 月 27 日
身份證字號	F129544372		
申請專案	Big Data Analysis Tools Development <input checked="" type="checkbox"/> 3 天(較佳) <input type="checkbox"/> 5 天		
導師或指導教授簽章	張啓文 11/8/2016	系所簽章	賴尚宏 Department of Computer Science National Tsing Hua University
通訊地址	台北市中山區堤頂大道二段 246 巷 3 號		
電話(Home)	(02)85027191(手機連絡)		
手機(Mobile)	0963243221(簡訊較佳)		
電子郵件	hubert052702@gmail.com		

# 2017 清華大學產業長期實習計畫



請按以下項目撰寫，字數不限，以完整清楚為要。

## 一、 列舉與申請實習項目相關的修課與成績：

### 大規模機器學習實務(研究所課程)，修課中

從理論開始教起機器學習的課程，同時兼顧實作所會遇上的問題與解決方法的課程，在寫這段文字的同時剛結束期中比賽 project：「根據新聞的 HTML 原始碼預測新聞的受歡迎程度」。我獨自用 Python 完成 preprocessing、feature 尋找與抽取、model 設計、training、tuning、testing 等課題，並應用 Boosting、Stacking、CrossValidation、Ensembling 等方法提升自己的準確率，即便比賽本身是分組比賽，但仍獨力在該比賽中取得班級排名第 2 的成績。

雖然在知識層面仍有需加強之處，但對於機器學習與資料科學已有初步的了解與認識，相信非常有助於在這個實習項目進入狀況。

### 資料科學概論(研究所課程)，修課中

與「大規模機器學習實務」不同，這門課較偏向理論性的介紹與概論，沒有對數學與實作深入探討，較偏向介紹 data science 的課題、會遇到的難題、可行的解決辦法與演算法，相信也是非常有助於在實習中快速進入狀況的經歷。

### 軟體實驗(Java)，A

關於 Java 的課程，final project 要寫一支 Android 程式，總行數 10000，雖不是非常有意義的指標，但能說明我對於 Java 語言有基本的理解與應用能力。

此外，在這個課程中與 project 的 server side 有使用到簡單的 SQL，讓我對 SQL 有最基本的了解，在實際實際使用上仍有不少需要學習之處。

## 二、 其他與實習項目相關學習經驗：

個人課外學習專注於 web 前後端與 javascript，與這個實習並無支接關係。

## 三、 實習動機：

近期接觸了不少資料科學與機器學習的課程，對於這領域相當有興趣，希望能脫離理論的紙上談兵，實際在會有使用者的狀況下實作專案、接觸問題、解決問題，也希望能實際在業界中學習軟體專案的管理/實作/測試/維護、向經驗豐富的前輩學習並改進自己的技術、藉由得到的經驗了解自己的能力不足之處並改進之。

#### 四、 對於實習的期待與想法：

希望能做出實際的成果與貢獻，就個人學習層面上得到實務經驗與自我改進的方向，就職涯的層面上提前理解自己未來要面對的問題、環境，實地應用自己過去所學的所有理論和知識解決實際問題，並嘗試拓展自己的人脈，最後，就最實際的實績層面上留下實績紀錄。

#### 五、 其他：

大三，實際上算是大四(降轉過)，差 18 學分畢業。

實習希望天數是填 3 天，主要是個人希望能留兩天修一門非常想修的課(雲端資料庫)和一門可能會影響到畢業的課，但若是實習需要(工作夠多之類的)，可以接受四天或五天。

# 林杰

# Hubert



> 清華大學資訊工程學系大三

> 語言：

Python/Java/

Javascript/C/C++

> 特殊專長：

MachineLearning/

Web 前端

> Github：

<https://github.com/hubert0527>

( 手機 ) 0963243221 ( 簡訊 )

hubert052702@gmail.com.tw

<https://hubert0527.github.io/>

## ➤ 應徵職位

Big Data Analysis Tools Development

3~5 天均可，但 3 天較佳，詳細安排希望能再面談。

## ➤ 自我介紹

名義上是大三，事實上是大學第四年，因曾就讀理學院所以平均 GPA 較低，資工課程平均 GPA 3.78 ( 略高於 A- )，實作能力屬於同儕前段，基本上所有期末/學期專案都能 ( 也經常是 ) 全部獨力實作，同時自我要求非常高，即便是獨力實作的團體作業也能得到高評價，基本上已擁有資工系畢業生應具備的基礎能力。

對於學習興趣濃厚，目前對於資料科學/機器學習深感興趣，因此正在修習兩門「研究所」課程，也在課堂上有優秀表現，此外也對軟體工程、網路前後端深感興趣，在課餘時間與寒暑假經常自行研究、實作個人專案。

這次申請實習希望能學習「軟體設計在業界的實際需求與要求」、「軟體測試」、「與他人協作經驗」、「累積大型程式實作經驗」，也期待自己的能力能夠回饋給公司、將交付的工作目標圓滿完成。

## ➤ 個人特質

非常擅長學習、挑戰新事物，並在學習歷程中會不斷思考、精進以符合自我評判的高標準。

喜歡解決問題、擅長除錯與尋找問題根源。

## ➤ 相關專案

### **News Popularity ( Large-Scale Machine Learning )**

相關連結：[Github](#)、[Kaggle](#):

這是「大規模機器學習」課程 ( 研究所課程 ) 的期中比賽，題目是根據新聞的 HTML 原始碼預測新聞的受歡迎度，此 project 為 3 人一組競賽，但實際由我「獨力」進行 preprocessing、feature 尋找與抽取、model 設計、training、tuning、testing 等所有工作，並依舊在該次競賽中得到第二名的成績 ( 修課人數 50 人 )。

從這個 project，我能證明自己對機器學習有基本的認識與實務能力，能使用 python 操作 scikit-learn 並多次自行實作演算法 ( Boosting、Stacking、CrossValidation、Ensembling )，雖然在知識層面仍有需加強之處，但對於機器學習與資料科學已有初步的了解與認識，相信非常有助於在這個實習項目進入狀況。

## *Data Crawling (MLB game information crawling)*

相關連結：[Github](#)

「資料科學概論」課程的一項作業，資料科學中很重要的一部分是取得資料與清理，我的目標是爬取歷年 MLB 所有比賽的數據。

我使用 python 的 request 與 selenium 分別進行一般頁面資料與 JS render 頁面資料的擷取，並針對所有可能的例外狀況進行錯誤處理、清理，實作結果抓取橫跨 103 年、17 萬場比賽、652 項 feature、共計 389MB、經過清理的 MLB 比賽資訊，對於資料擷取與清理有更深入與實務上的經驗。

## *CoWalkr ( Software Studio )*

相關連結：[Github](#)

軟體實驗課程的 Android 程式 final project，總行數 10000 行，雖然不是非常有意義的指標，但能證明我對 Java 有一定程度上的理解與熟悉。在此專案中，我作為組長，對程式的運作流程進行規劃、實作個別頁面的雛型與許多細部功能實作，並盡力控制專案的開發時間。

此外，在 server side 實作中有使用到 SQL query 資料，對於 SQL 有最基本的了解，實際使用上仍有不少需要學習之處。

## *MIPS ISA3000 Simulator ( Computer Architecture )*

相關連結：[Github](#)

實習條件中有提及 C 語言，所以這裡提出自己目前認為最符合自己 C 語言水準的專案，約 3000 行，有簡單的資料結構與經規劃的函式切分大幅增加程式的可讀性，並沒有使用 OOP。

此專案要對計算機結構、記憶體管理、組語等項目有深入理解，並以 C 語言實作模擬 MIPS CPU 處理組合語言的行為，在課程中需經歷三次重構，考驗程式的可重複使用性、結構完整性。

此外，課程中經教授同意，額外提供 binary/assembly 雙向轉譯程式與大量測資測試程式供同儕使用（皆是 C 語言實作），最終此專案在課程中的 196 筆嚴格隱藏測資中通過 194 筆，通過率 99%，也在該課程中得到 A+ 評定。

## **其他專案**

其餘專案與此實習均無直接相關，不過都有在 github 上建立 project 可檢閱。（<https://github.com/hubert0527>）

# 自傳

## 學習程式歷程

我從開始學寫程式至今已經三年，實際上真的認真投入學習中是近兩年的事，原先就讀的是理學院學士班，第一專長是物理，第二專長數學，而後因發覺興趣轉變而轉到資工系。

因為有這一次的重來機會，所以深刻體悟到自己應把握這個機會，於是加倍認真投入課程。我在課內對於課程有相當深入的了解，並在各項課程 **project** 中都以高標準自我要求，在實技課程中都獲得教授 **A** 或 **A+** 的肯定。

此外，碰巧有幸認識高中同校的一位朋友，看到他在課程之外接了許多成果令人驚豔的案子與他的學習歷程，受到非常強烈的震撼。因此我在課餘之時也非常積極於自我學習，讓我比起一班資工系學生的見聞更加廣闊，但這部分的學習重點較專注於網路前後端的學習，與此次的實習並無太大關係，所以並沒有在履歷中特別詳述。

## 興趣與實技

興趣方面，個人對於機器學習、數據分析、大型軟體專案開發與網路前後端設計都非常有興趣，所以不顧過大的選課壓力而在這學期選修了兩門研究所關於機器學習與大數據的課程，在兩門課程中也都有非常多的收穫，並希望能更進一步了解與學習這些技術在業界中是如何被應用的，於是申請這次實習。

實技方面，因為我是轉系生，認識的人較少，所以幸也不幸，大多數狀況下的分組作業都要面臨與分組分剩的學生同組（近期因為在課堂上有一些傑出表現，開始會遇上志同道合的同學了）。不幸之處是比較操勞，而幸運之處在於我更加珍惜、把握能與組員一同合作的機會、努力超越教授設定的目標。在必須獨自面對多人專案時，我也不畏懼挑戰，盡力獨自做出比多人合作更好的成果，透徹、全面地瞭解教授透過專案希望傳達的概念。

## 期許

這次申請實習，一方面是希望能增廣見聞、增加實技經驗、了解自己所學的知識在產業界是如何應用的。另一方面，在學校分工合作時也經常有感自己的帶領能力有所不足，希望能學習如何在工作上與團隊分工合作，並觀察前輩們是如何溝通、帶領團隊、劃分工作職責、規劃時程，並增進自己的與團隊合作的能力。