

Contents

1	Globalisation	3
1.1	Towards globalisation	3
1.2	Internationalisation (i18n)	3
1.3	Localisation (l10n)	4
1.4	Globalisation (g11n)	5
2	Text encoding	7
2.1	A brief history of character encoding	7
2.1.1	Bi-literal cipher of Francis Bacon	7
2.1.2	Gauss and Weber's code	8
2.1.3	Morse and Gerke's code	8
2.1.4	Baudot code	9

Chapter 1

Globalisation

1.1 Towards globalisation

An ultimate goal of any for-profit organisation is making money. There are two extremes to achieve it: sell limited luxury goods or services at a very high price, or sell something to the mass market at a low price. In the software realm, the latter usually is taken as a strategy. At first, a startup is created to test an idea, and then the product evolves to something bigger, reaches more markets and countries until almost any person in the world can use it. In business, the last step is called "going global", or simply "globalisation".

To achieve *globalisation* for a software product, a company, at first, should accomplish product *internationalisation* and *localisation*. There are no strict definitions for the italicized terms these days, and people treat the notions differently ([1] [2] [3] [4]). In the next sections common definitions are provided for these terms.

1.2 Internationalisation (i18n)

Software internationalisation implies that the product is functional with any language input and has easy localisation.

Internationalisation is also known as *i18n*, where 18 is the number of letters between i and n. This shortened form is called *numeronym* [5]

Localisation is shortened to *l10n* (1 + 10 letters + n), and globalisation is written as *g11n* respectively.

Internationalisation is often reached by the following means:

- Support of various human language symbols.

Regularly UTF-8 is used as a text encoding standard for display and input symbols. A Japanese person should be able to write こんにちは (Hello) to a friend living in America, and the recipient should be able to read the letter without any additional action.

- Acceptance of world numeric, date, time, and currency formats.

Having default regional settings, Chilean customers should be able to put "1,5" value to the system, while Australian users should see it as "1.5" automatically.

- Easy addition of UI localization resources.

Netflix operates in more than 190 countries, and for UI localisation created the product called Hydra [**hydranetflix**]. Java has the ResourceBundle class for new languages. Similarly, Microsoft provides a fast and easy method to add new languages for .NET projects.

1.3 Localisation (l10n)

Localisation is the process of a product adaptation to meet the language and cultural requirements of a specific target market.

To get localisation done, companies often adopt the following practices:

- Translation of messages and texts to the localised version.
- Verification of the layout correctness for the targeted language.

Chinese words regularly take less space than German words and the UI interface should be flexible to support both languages.

- Conforming to local standards.

In 2016, the EU adopted the General Data Protection Regulation (GDPR) to protect EU citizens data privacy [6]. In two years later, the Barreiro Hospital in Portugal was fined 400,000 € for providing data access to many former employees [7]. One more GDPR requirement is to advise users about private data usage that resulted in the annoying pervasive popup "we use browser cookies to watch you for the good".

In the same year, the Russian government signed the "Russian GDPR" as part of the "Yarovaya Law" [8] [9]. Microsoft and Facebook moved personal citizens data to their in-state data centres, while LinkedIn

was not ready for such changes and the government blocked the web resource.

Another example involves tax support for a new region. The sales tax in the USA varies from state to state. In contrast, in Germany, the value-added tax (VAT) rate is the same for the whole country, and the calculation logic is much different. This is only a small example of a distinction between the two countries. It took several years for SAP, a Germany company, to adopt the software for the the US market and be among the top enterprise systems.

1.4 Globalisation (g11n)

Globalisation is the process by which organizations connect with their customers and partners around the world.

Software globalisation often realized in:

- Neutral attitude to every presentation aspect.

It is not always possible to foresee all meanings of the word in various languages. Mitsubishi had to rename Pajero vehicle line to Montero for Spanish version because of the offensive sense. Likewise, Hyundai renamed the Kona model to Kauai on the Portuguese market.

- Software design, development, and customers support are being presented at least in every continent.

Google has offices in more than 50 countries [10]. Microsoft offices are presented in almost every country [11]. Every global player has a sales and a support team in every large city (see [12, Apple offices], [13, SAP offices], [14, Amazon offices]).

In a simple form, at the internationalisation step, a plan and a platform is prepared for localisation, periodically the plan and the methods are updated, the product is localised and reaches globalisation:

Internationalisation → Localisation → Globalisation

Chapter 2

Text encoding

2.1 A brief history of character encoding

2.1.1 Bi-literal cipher of Francis Bacon

In 1605, Francis Bacon, an English philosopher, scientist and statesman, devised the cipher that corresponds the English alphabet to a 5-bit code [15].

Letter	Code	Binary
A	aaaaa	00000
B	aaaab	00001
C	aaaba	00010

F	aabab	00101

L	ababa	01010

Y	babba	10110
Z	babbb	10111

Table 2.1: Bacon's cipher

The most powerful insight was that this cipher can be used for encoding by any two types of state: a pair of font style, "as by Bells, by Trumpets, by Lights and Torches, by the report of Muskets, and any instruments of like nature".

Below is the example provided by Bacon. If a bold font style denotes "a" and a normal font style denotes "b", then the phrase of 15 letters "Do not

go til I come” can be transformed to:

Do not go til I come

↓

aababababababba

↓ (group by 5)

aabab ababa babba

↓ (decode)

F L Y

In 1916, William Frederick Friedman, a US Army cryptographer, created an illustration how ”anything can signify anything” with the Bacon’s cipher [16].

2.1.2 Gauss and Weber’s code

In 1833, Carl Friedrich Gauss and Wilhelm Weber, two German scientists, succeeded in transmitting messages via an electromagnetic data line at Göttingen Observatory. Although the invented telegraph had no financial funding and no further development, the variable 5-bit character code was invented [martin2010technological].

2.1.3 Morse and Gerke’s code

Samuel Morse invented the code for English alphabet, numbers and basic punctuations. In 1844, he send the first public American telegram from Washington to Ohio with the message ”what hath God wrought?” (”what has God done?”). In the first version of Morse code each symbol is represented by a sequence of pulses of 4 length variants (base 4).

In 1848, Friedrich Gerke, a German writer, musician and telegraph technician, revised Morse’s code. The new simplified version had a sequence of only a short and a long signal variants (base 2). Later, Gerke’s code was adopted to International Morse-Gerke Code.

From the encoding perspective, we have the same pair of bits, now presented by the signal duration.

As an example, consider encoding of the famous emergency signal SOS that was introduced in the beginning of 20th century.

S	O	S
...	- - -	...
000	111	000
3 short	3 long	3 short

Table 2.2: SOS code

2.1.4 Baudot code

In 1870s, Émile Baudot, a French telegraph engineer, invented a multiplex system for the telegraph with the new 5-bit code for Roman alphabet, punctuation and control signals. The code was used in telegraph prior the ASCII

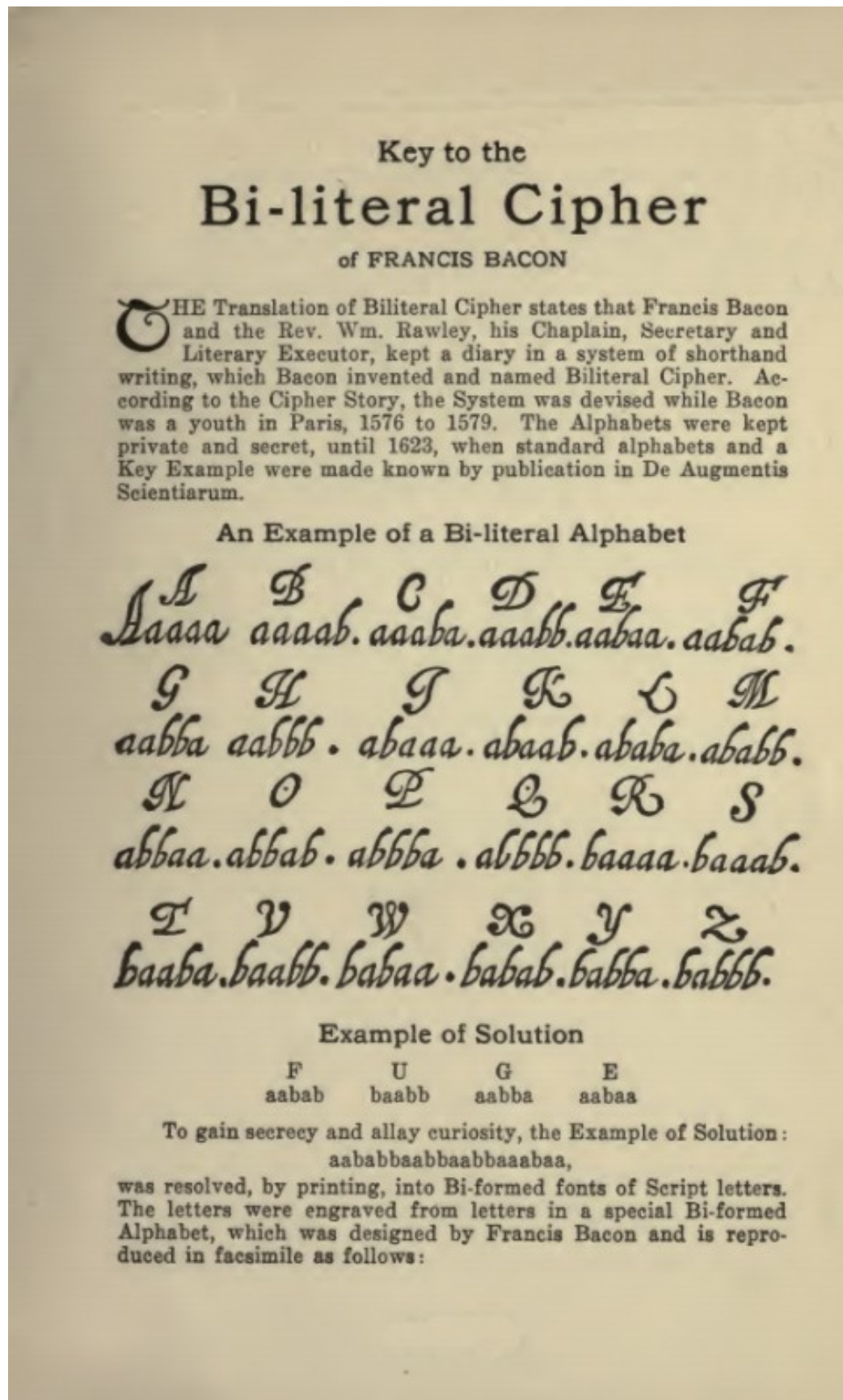


Figure 2.1: Francis Bacon, *Of the Proficiency and Advancement of Learning, Divine and Human*, 1605

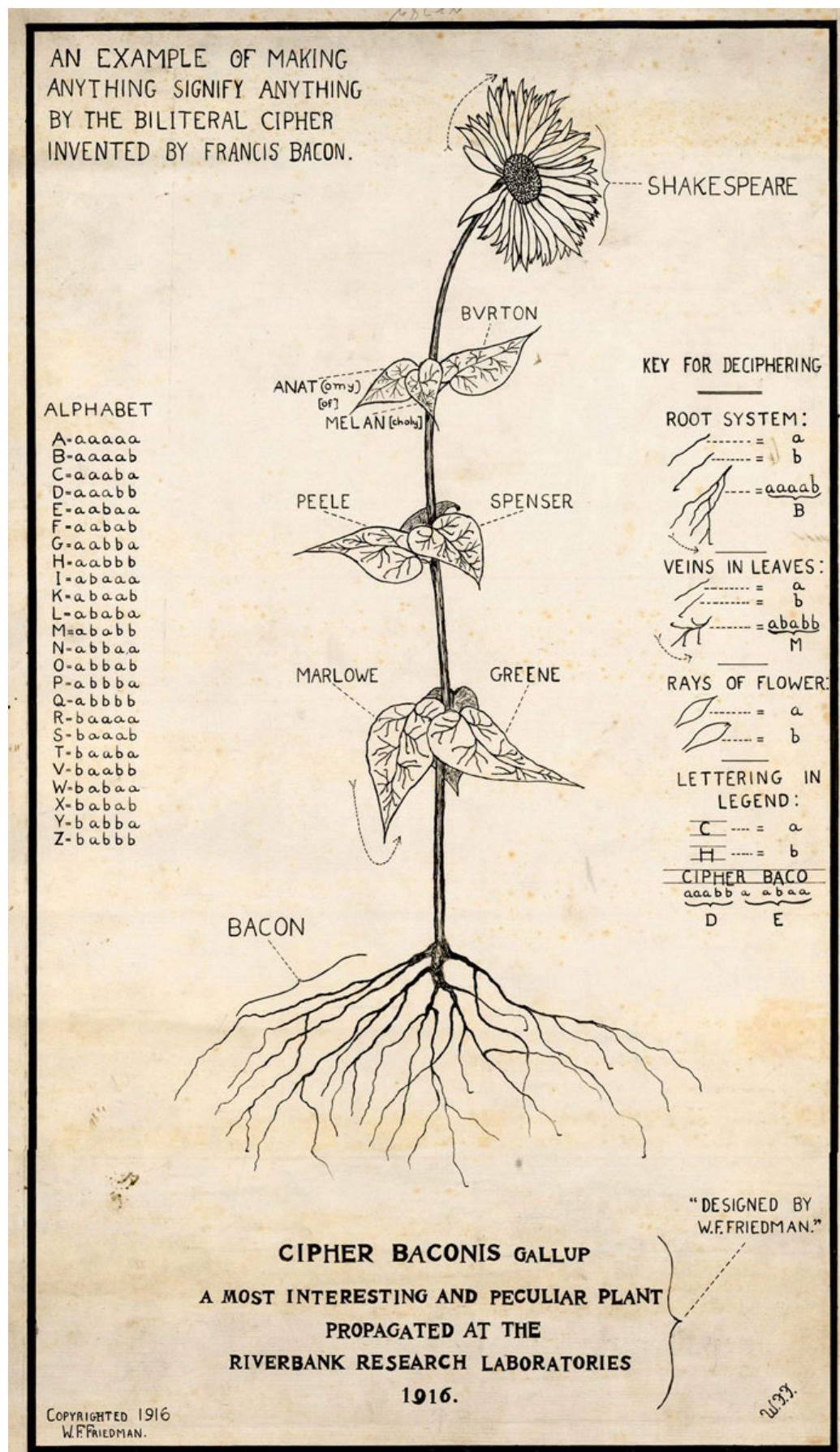


Figure 2.2: W.F. Friedman, Cipher Baconis Gallup, 1916

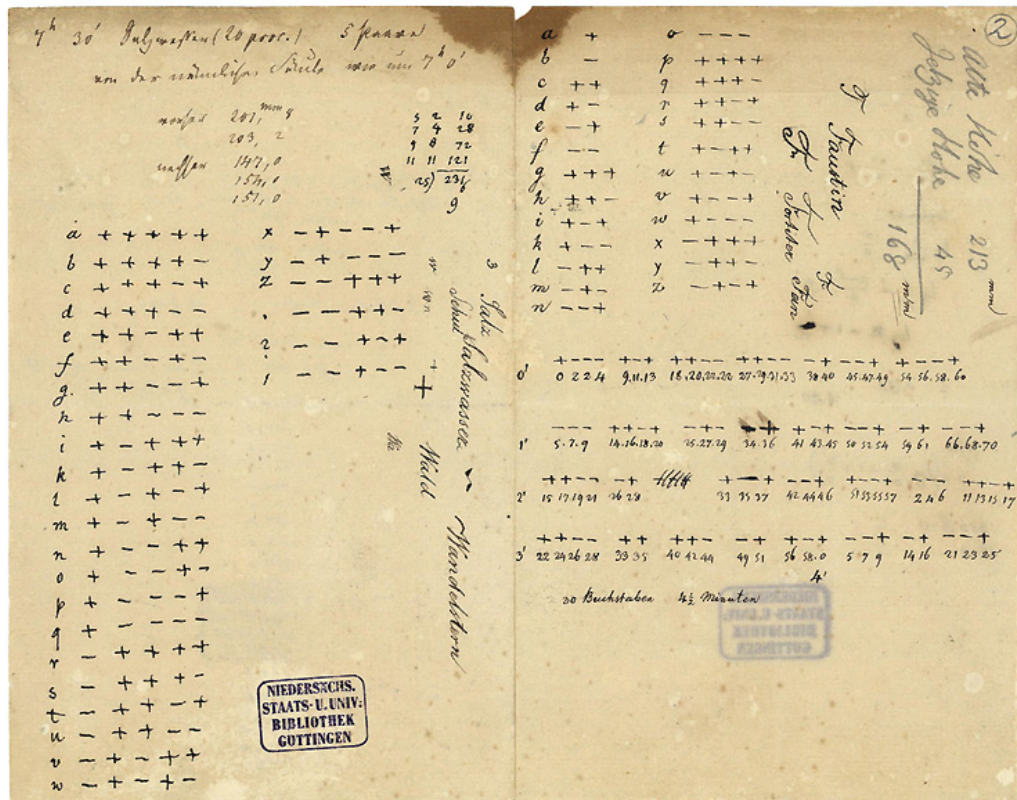


Figure 2.3: Gauss and Webber's code, Göttingen University

	American (Morse)	Continental (Gerke)	International (ITU)
A	• —	• —	• —
Ä		• — • —	
B	— • • •	— • • —	— • • •
C	• • •	— • — •	— • — •
CH		— — — —	
D	— • •	— • •	— • •
E	•	•	•
F	• — •	• • — •	• • — •
G	— • — •	— — — •	— — — •
H	• • • •	• • • •	• • • •
I	• •	• •	• •
J	— • • — •	• — — — —	• — — — —
K	— • • —	— • — —	— • — —
L	— — —	• — — •	• — — •
M	— — —	— — —	— — —
N	— •	— •	— •
O	• •	• — • • •	— — — —
Ö		— — — — •	
P	• • • • •	• • • • •	• — — — •
Q	• • — • •	— — — • —	— — — • —
R	• • • •	• — — •	• — — •
S	• • •	• • •	• • •
T	—	—	—
U	• • —	• • —	• • —
Ü		• • — — —	
V	• • • —	• • — —	• • • —
W	— — —	— — —	— — —
X	• • • •	• • — — • •	— • • — —
Y	• • • •	— — — • • •	— — — — —
Z	• • • •	• — — — • •	— — — • •
1	• — — — •	• — — — •	• — — — —
2	• • — — • •	• • — — • •	• • — — — —
3	• • • — • •	• • • — • •	• • • — — —
4	• • • • —	• • • • —	• • • • —
5	— — — —	— — — —	• • • • •
6	• • • • • •	• • • • • •	— • • • • •
7	— — — • •	— — — • •	— — — • • •
8	— • • • • •	— • • • • •	— — — — • •
9	— • • • •	— • • • •	— — — — — •
0	— — — — —	— — — — —	— — — — —
0 (alt)	—		—

Figure 2.4: Morse, Gerke, ICU codes, Courtesy Spinningspark at Wikipedia

Bibliography

- [1] Dimitra Anastasiou and Reinhard Schäler. “Translating vital information: Localisation, internationalisation, and globalisation”. In: *Syntheses Journal* 3 (2010), pp. 11–25.
- [2] Susan K. Miller Richard Ishida. *W3C Internationalisation Q&A*. 2005. URL: <https://www.w3.org/International/questions/qa-i18n> (visited on 08/17/2019).
- [3] Michael S. Kaplan. *Internationalization vs. localizability, MSDN blog archive*. 2005. URL: <http://archives.miloush.net/michkap/archive/2005/09/09/462862.html> (visited on 08/17/2019).
- [4] Microsoft. *Globalizing and localizing .NET applications, .NET Guide*. 2019. URL: <https://docs.microsoft.com/en-us/dotnet/standard/globalization-localization/index> (visited on 08/17/2019).
- [5] Tex Texin. *Origin Of The Abbreviation I18n*. URL: <http://www.i18n.guy.com/origini18n.html> (visited on 08/17/2019).
- [6] *GDPR*. 2016. URL: <https://eugdpr.org> (visited on 08/17/2019).
- [7] Datenschutz notizen. *Portuguese Data Protection Authority Imposes 400,000 Eur Fine on Hospital*. 2018. URL: <https://www.datenschutz-notizen.de/portuguese-data-protection-authority-imposes-400000-e-fine-on-hospital-4821441/> (visited on 08/17/2019).
- [8] International Law Office. *Yarovaya Law and new telecoms data storage requirements*. 2018. URL: http://www.gorodissky.com/upload/articles/files/Yarovaya_Law_and_new_telecoms_data_storage_requirements.pdf (visited on 08/17/2019).
- [9] ICNL. *Overview of the Package of Changes into a Number of Laws of the Russian Federation Designed to Provide for Additional Measures to Counteract Terrorism*. 2016. URL: <http://www.icnl.org/research/library/files/Russia/Yarovaya.pdf> (visited on 08/17/2019).
- [10] Google. *Google office locations*. 2019. URL: <https://careers.google.com/locations/> (visited on 08/17/2019).

- [11] Microsoft. *Microsoft office locations around the world*. 2019. URL: <https://www.microsoft.com/en-us/worldwide.aspx> (visited on 08/17/2019).
- [12] Apple. *Apple find locations*. 2019. URL: <https://locate.apple.com/> (visited on 08/17/2019).
- [13] SAP. *SAP Worldwide directory*. 2019. URL: <https://www.sap.com/corporate/en/company/office-locations.html> (visited on 08/17/2019).
- [14] Amazon. *Amazon Locations*. 2019. URL: <https://www.amazon.jobs/en-gb/locations/> (visited on 08/17/2019).
- [15] Francis Bacon. *Of the Proficiency and Advancement of Learning, Divine and Human*. 1605. URL: <https://archive.org/details/advancementof100baco/page/256> (visited on 08/17/2019).
- [16] W.F. Friedman. *Cipher Baconis Gallup*. 1916. URL: <https://digitalcollections.nypl.org/collections/bacon-cipher-collection> (visited on 08/17/2019).