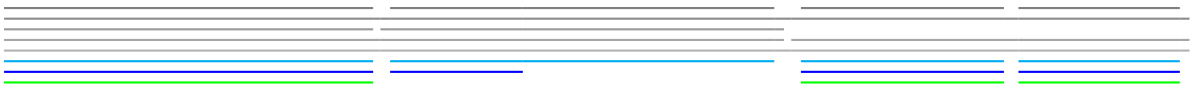
A Appendix

We present below the annotations of 8 short biographies of 20th century scientists according to 5 human annotators, C-sanitize, the neural NER model and the Presidio anonymisation tool (see paper for details). The annotation task[1] consisted of tagging text spans that could re-identify a person either directly or in combination with publicly available knowledge. The annotators were instructed to prevent identity disclosure, but otherwise seek to preserve the semantic content as much as possible. The five annotators were researchers in statistics and natural language processing.

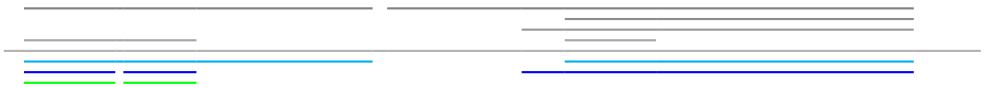
The first five (gray) lines denotes the five human annotators, while the cyan line corresponds to C-sanitise, the blue line to the neural NER model, and the green line to the Presidio tool.

A.1 Alexander Frumkin

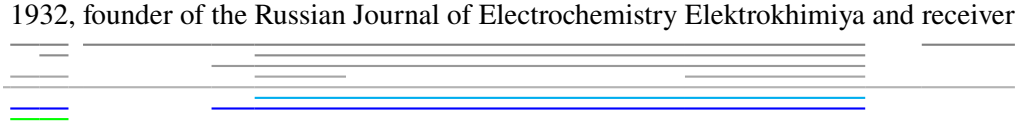
Alexander Naumovich Frumkin (Александр Наумович Фрумкин) (October 24, 1895–May 27, 1976)



was a Russian/Soviet electrochemist, member of the Russian Academy of Sciences since

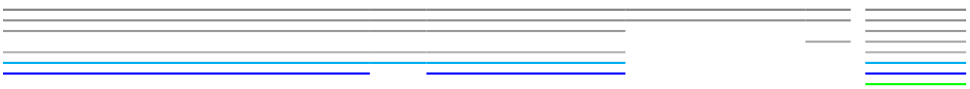


1932, founder of the Russian Journal of Electrochemistry Elektrokhimiya and receiver

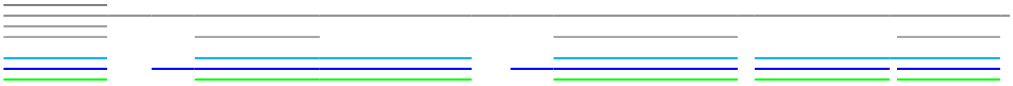


of the Hero of Socialist Labor award. The Russian Academy of Sciences' A. N. Frumkin

Institute of Physical Chemistry and Electrochemistry is named after him. Frumkin was



born in Kishinev, in the Bessarabia Governorate of the Russian Empire (present-day Moldova)



to a Jewish family; his father was an insurance salesman. His family moved to Odessa,

where he received his primary schooling; he continued his education in Strasbourg, and

then at the University of Bern. Frumkin's first published articles appeared in 1914,

[1]The guidelines and annotated data are publicly available: https://github.com/anonymous-NLP/anonymisation

when he was only 19; in 1915, he received his first degree, back in Odessa. Two years later, the seminal article “Electrocapillary Phenomena and Electrode Potentials” was published. Frumkin moved to Moscow in 1922 to work at the Karpov Institute, under A. N. Bakh. In 1930 Frumkin joined the faculty of Moscow University, where in 1933 he founded—and would head until his death—the department of electrochemistry. During the Second World War, Frumkin led a large team of scientists and engineers involved in defense issues. This contribution did not save him from being dismissed in 1949 as the director of the Institute of Physical Chemistry, when he was accused of “cosmopolitanism”. Frumkin’s most fundamental achievement was the fundamental theory of electrode reactions, which describes the influence of the structure of the interface between electrode and solution on the rate of electron transfer. This theory has been confirmed and extended within the framework of contemporary physical electron transfer models. Frumkin introduced the concept of the zero charge potential, the most important characteristic of a metal surface.

Alessandro Volta's question—a topic of discussion for over 120 years—about the nature of the EMF of electrochemical circuits was resolved using Frumkin's approach. Frumkin developed the Frumkin isotherm, an extension of the Langmuir isotherm in describing certain adsorption phenomena. Frumkin's students developed novel experimental methods that would, in time, become standard. Several applied electrochemical processes, including ones related to chemical sources of electrical power, industrial electrolysis, and anti-corrosion protection, were successfully developed under Frumkin's supervision. Frumkin was married three times, including a brief first marriage to Vera Inber.

A.2 Bashir Rameev

Bashir Iskandarovich Rameev (Russian: Башир Искандарович Рамеев; formerly "Rameyev" in English; 1 May 1918–16 May 1994) was a Soviet inventor and scientist, one of the founders of Soviet computing, author of 23 patents, including the first patent in the field of electronic computers officially registered in the USSR—a patent for the Automatic Electronic Digital Machine (1948). Rameev's inventions paved the way for the development of a new

field in Soviet science—electronic computing—and for the formation of a new branch of industry that supported it. The central ideas incorporated in Rameev's invention of the electronic computer included: storing programs in computer memory, using binary code, utilizing external devices, and deploying electronic circuits and semiconductor diodes.

The first publication about similar technology outside of the USSR appeared in 1949-1950.

Rameev also suggested that intermediate computation data be automatically printed on punched tape and sent into the computer's arithmetic device for subsequent processing, meaning that the processing of commands would be performed in the computer's arithmetic device; this is usually referred to as the Von Neumann architecture. Of particular note is Rameev's invention of diode-matrix control circuits, which were used to build his first brainchild, the first serially manufactured Soviet mainframe "Strela" (1954).

In the 1950s, the diode-matrix control circuits were not widespread due to their significant dimensions and high power consumption. However, with subsequent development of microelectronics and the emergence of large-scale integrated circuits, which made possible to deploy tens

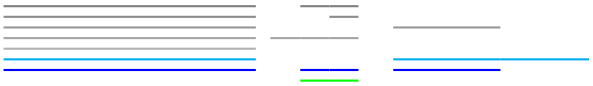
or hundreds of thousands of diodes and transistors in a single piece of silicon, the concept of control circuits became viable and commonly used. “Strela” computers carried out calculations in nuclear physics, rocketry and space research. Notably, one of “Strelas” was used to calculate “Sputnik” orbit trajectory. For the development of “Strela” Rameev and his team were awarded the Stalin Prize of 1st degree, which was the highest Soviet award at that time. Between 1956 and 1969, Rameev designed and oversaw the manufacturing of 14 different computers including: the multi-purpose “Ural” computer series and the specialized machines “Weather” (“Погода”), “Crystal” (“Кристалл”), “Granite” (“Гранит”), and “Coordinate” (“Координата”). Rameev’s “famous computer family ’Ural’ existed more than 15 years and had good chances to be one of the corner stones of future Russian computer engineering”.

A.3 Brian Mcconaghy

Brian McConaghy (born 1963) is the founder of Ratanak International and a former Canadian forensic scientist who left the Royal Canadian Mounted Police (RCMP) in order to dedicate

all his energies to ending child abuse and human trafficking in Cambodia. He had already

founded Ratanak International, in 1989, a Christian charity dedicated to helping the



people of Cambodia rebuild their country that for decades had been torn apart by civil

war, revolution and genocide. From 1990 onwards McConaghy and Ratanak partnered on projects

that built clinics, hospitals and schools, opened orphanages, provided shelters for the

elderly and AIDS victims and ran and initiated emergency food distribution programs in

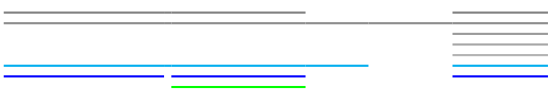
response to droughts and flooding in Cambodia. In 2004, these relief projects continued,

yet Ratanak's work also took on a whole new dimension by beginning to work on the front

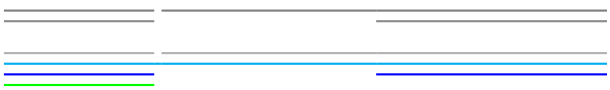
lines in Cambodia on projects that rescue and rehabilitate children sold into sexual

slavery. McConaghy named the organization Ratanak, which means 'precious gem' in Khmer,

after he watched an 11-month-old Cambodian baby called Ratanak die because of a basic



lack of medical aid in John Pilger's documentary film Cambodia Year Ten. Since watching



this video McConaghy and Ratanak have been dedicated to preventing such needless suffering

and death in Cambodia. McConaghy grew up in Northern Ireland and his family emigrated to Canada in 1978. He used to work for the Vancouver Forensic Laboratory as a firearm and tool-mark examination specialist. After founding Ratanak International, he continued to help the RCMP as a consultant, providing them with crucial information about the identity of the child sexual abuse victims in the case of Donald Bakker. He also did forensic work on the women murdered by Robert Pickton, testifying that Andrea Joesbury, Sereena Abotsway, and Mona Wilson had all been decapitated with a reciprocating saw. McConaghy was the guest speaker at the 2011 Manitoba Prayer Breakfast. He lives in Vancouver with his wife and two children, both adopted from Cambodia.

A.4 Freeman Dyson

Freeman John Dyson (15 December 1923–28 February 2020) was a British-American theoretical and mathematical physicist, mathematician and statistician known for his works in quantum field theory, astrophysics, random matrices, mathematical formulation of quantum mechanics, condensed matter physics, nuclear physics and engineering. He was professor emeritus

in the Institute for Advanced Study in Princeton, a member of the Board of Visitors of Ralston College and a member of the Board of Sponsors of the Bulletin of the Atomic Scientists.

Dyson originated several concepts that bear his name, such as Dyson's transform, a fundamental technique in additive number theory, which he developed as part of his proof of Mann's theorem; the Dyson tree, a hypothetical genetically-engineered plant capable of growing in a comet; the Dyson series, a perturbative series where each term is represented by Feynman diagrams; the Dyson sphere, a thought experiment that attempts to explain how a space-faring civilization would meet its energy requirements with a hypothetical megastructure that completely encompasses a star and captures a large percentage of its power output; and Dyson's eternal intelligence, a means by which an immortal society of intelligent beings in an open universe could escape the prospect of the heat death of the universe by extending subjective time to infinity while expending only a finite amount of energy.

Dyson disagreed with the establishment scientific position that carbon dioxide (CO2) is a material driver of planetary temperature increases. He believed that some of the

effects of increased CO2 levels are favourable and not taken into account by climate scientists, such as increased agricultural yield and further that the positive benefits of CO2 likely outweigh the negative effects. He was skeptical about the simulation models used to predict climate change, arguing that political efforts to reduce causes of climate change distract from other global problems that should take priority. He also signed the World Climate Declaration titled "There is no Climate Emergency".

A.5 Khudoyor Yusufbekov

Khudoyor Yusufbekovich Yusufbekov (Russian: Худоер Юсуфбекович Юсуфбеков, Tajik: Худоёр Юсуфбеков; December 10, 1928—November 27, 1990) was a Soviet scientist and organizer of scientific projects and institutes in Pamir. He was a leading scientist who made a significant contribution to the development of biological sciences, whose name is connected with a new direction of the development of plant growing in the arid mountain and highland territory of Pamir-Alay; a prominent specialist in the field of plant growing, plant introduction and pasture economy, meadow studies, phyto-amelioration, and botany, Yusufbekov

was a practicing field researcher, figure of higher education, and professor. In 1968,

he developed a system for fodder improvement in the Pamir and Alay valleys that was differentiated

from the perspective of the ecological and geographical areas and high-altitude zones.

He also implemented a system of arid fodder, and proposed methods of cultivation of useful

plants in the Pamir area in 1972. In 1970—1975, Khudoyor Yusufbekov developed the master

plan of reconstruction of the Pamir Botanical Garden. In 1969, he became doctor of the

agricultural sciences. In 1976, he became an Academician of the Academy of Sciences of

the Tajik Soviet Socialist Republic. In 1962—1969, he was the director of the Pamir Biological

Station; at the same time in 1965—1990, he was the Chairman of the Bureau of the Pamir

Base; in 1969—1981, the director of the Pamir Biological Institute of the Academy of

Sciences of the Tajik SSR; in 1981—1986, the rector of the Tajik Agricultural Institute

of the Ministry of Agriculture of the USSR; in 1986—1990, the Academician Secretary of

the Biological Department of the Academy of Sciences of the Tajik SSR. From 1989, he

was a Member of the Presidium of Academy of Sciences of the Tajik SSR. Moreover, he was a state and public figure, the head of the scientific council of the department of biological science of the Academy of Sciences of the Tajik SSR and a Member of the coordination council of the department of general biology of the Academy of Sciences of the USSR (1987—1990). He was also a fellow of the Geographical Society of the USSR since 1965, Member of the All-Union and Central Asian Councils of the Botanical Gardens of the USSR (1972—1990), Member of the Council on the "Biological Foundations of the Rational Use and Protection of Flora" of the Academy of Sciences of the USSR (1976—1990), Member of the Council on the "Biological Foundations of the Development of Mountain Territories in Central Asia" (1975—1990), Member of the Council of the All-Union Botanical Society (1976—1990).

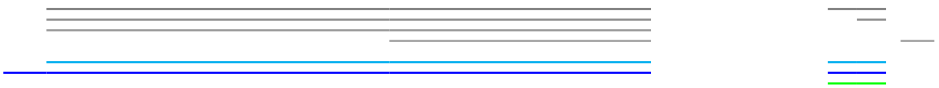
A.6 Oswaldo Frota-Pessoa

Oswaldo Frota-Pessoa (March 30, 1917–March 24, 2010) was a noted Brazilian physician, biologist and geneticist. Oswaldo Frota-Pessoa was born in Rio de Janeiro, where he did

all his studies, first in natural history at the Federal District University (currently

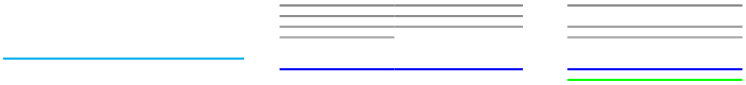
the State University of Rio de Janeiro), graduating in 1938; and subsequently medicine

at the National School of Medicine of University of Brazil, graduating in 1941. He got



his doctoral degree at the same school, in 1953 and soon afterwards went abroad on a

scholarship for post-doctoral studies at Columbia University, in New York City, from



1953 to 1955. His teaching and research professional career began in 1942, when he accepted

a position of assistant professor at the School of Philosophy, Sciences and Letters of

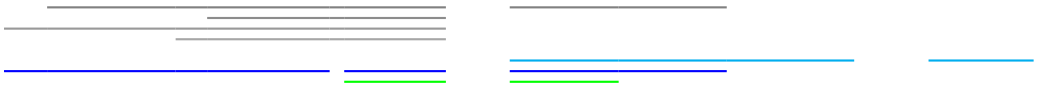
the Federal University of Rio de Janeiro, a post he held until 1958. In that year he

accepted a new position at the University of São Paulo, moving to São Paulo City, where

he worked until his retirement. He attained a full professorship there in 1978 and was

elected an emeritus professor in 1995. In 1964 and 1965 he was a Visiting Professor at

the University of Wisconsin–Madison on a Fulbright Program fellowship. As a scientific



leader, Dr. Frota-Pessoa held many prominent positions, such as: specialist in Science

Education of the Pan American Union (Organization of American States) in Washington, D. C. (1955–1956), consultant in Human Genetics for the World Health Organization (1961–1986), director of the Coordination Center of Brazil of the Multinational Program of Genetics of the Pan American Union (1968–1973), director of the Centro de Estudos sobre Currículo para o Ensino de Biologia (CECEB) from 1972 to 1979, president of the Brazilian Society of Genetics (1968–1970) and of the Latin American Association of Genetics (1969–1971), founding member of the Academy of Sciences of the State of São Paulo (1974). He published more than 130 research papers on genetics and about 500 popularization articles. His main research interests were the systematics of Drosophila, the genetics of human populations, cytogenetics, medical genetics and genetic counseling, and genetics in psychiatry. Dr. Frota-Pessoa was always one of the most active and respected enthusiasts for the teaching of biology and the popularization of science and a promoter of public understanding of science. He actually taught science and biology in secondary schools of the public system of Rio de Janeiro from 1939 to 1958. Based on this experience, he wrote one of the first

textbooks on biology for secondary education, which became a best-seller and was published in many editions. In all, he published 26 textbooks and 17 guides for science and biology teachers. For these efforts, he won the UNESCO Kalinga Prize for the Popularization of Science and the CNPq José Reis Award for the Divulgation of Science. He was also decorated by the Brazilian government with the Great Cross of the Brazilian Order of Scientific Merit and was awarded the 1989 Alfred Jurzikowyski Prize of the Brazilian Academy of Medicine, for relevant basic research for medicine.

A.7 Richard Feynman

Richard Phillips Feynman (; May 11, 1918–February 15, 1988) was an American theoretical physicist, known for his work in the path integral formulation of quantum mechanics, the theory of quantum electrodynamics, the physics of the superfluidity of supercooled liquid helium, as well as his work in particle physics for which he proposed the parton model. For contributions to the development of quantum electrodynamics, Feynman received the Nobel Prize in Physics in 1965 jointly with Julian Schwinger and Shin'ichirō Tomonaga.

Feynman developed a widely used pictorial representation scheme for the mathematical

expressions describing the behavior of subatomic particles, which later became known

as Feynman diagrams. During his lifetime, Feynman became one of the best-known scientists

in the world. In a 1999 poll of 130 leading physicists worldwide by the British journal

Physics World, he was ranked as one of the ten greatest physicists of all time. He assisted

in the development of the atomic bomb during World War II and became known to a wide

public in the 1980s as a member of the Rogers Commission, the panel that investigated

the Space Shuttle Challenger disaster. Along with his work in theoretical physics, Feynman

has been credited with pioneering the field of quantum computing and introducing the

concept of nanotechnology. He held the Richard C. Tolman professorship in theoretical

physics at the California Institute of Technology. Feynman was a keen popularizer of

physics through both books and lectures, including a 1959 talk on top-down nanotechnology

called There's Plenty of Room at the Bottom and the three-volume publication of his undergraduate

lectures, The Feynman Lectures on Physics. Feynman also became known through his semi-autobiographical

books Surely You're Joking, Mr. Feynman! and What Do You Care What Other People Think?,

and books written about him such as Tuva or Bust! by Ralph Leighton and the biography

Genius: The Life and Science of Richard Feynman by James Gleick.

A.8 Vladislav Ivanov (Physicist)

Vladislav Alexandrovich Ivanov (Russian: Владислав Александрович Иванов; 1936-2007) was

a Soviet physicist and engineer, who proposed in 1959 the basic principles of Magnetic

Resonance Imaging, decades before this technique was demonstrated by Paul Lauterbur.

Ivanov graduated from Lenigrad Airforce Academy in 1959. While at the academy, he came

up with the idea of using the recently discovered phenomenon of Nuclear Magnetic Resonance

for imaging purposes. In 1959, he filed his first application for Invention Certificate

(a patent-like document used in the Soviet Union) titled "Free-precession proton microscope".

Soon afterwards he filed three more applications. The second of his application (filed

in March of 1960) comprised a detailed description of the MRI principles, as was confirmed

more recently. Originally this application was rejected as “unrealizable”. However,

in 1984 an Invention Certificate № 1112266‘‘ A method for determination of internal structure

of material objects” was finally issued in 1984, only after this method was demonstrated

in other countries. After leaving the military, Ivanov returned to Leningrad, where he

enrolled in Saint Petersburg Electrotechnical University, which he graduated from in

1966 with a PhD in Engineering. In 1967 he became a lab director at “Elektroavtomatika”

design bureau, and in 1969 a lab director at D. I. Mendeleyev Institute for Metrology

(VNIIM). In 1980 he received his habilitation, and 1984 he was promoted to the rank of

professor at ITMO University. Despite his failure to commercialize his MRI invention,

Ivanov continued his career as a prolific inventor. His name is listed on over 100 patents.

He was a developer of apparatuses for space, aviation, marine and underground applications.

He was the lead designer of two Soviet National standards: of angular velocity and of

acceleration. Ivanov was a member of the American Mathematical Society and the International Society of Automation. Ivanov wrote over 300 books and articles, including 3 books of poetry (published in 1991, 1997, and 1999).