

ДЕЯКІ РОЗДУМИ ПРО ТЕ, ЯК МУСИТЬ БУТИ ВЛАШТОВАНИЙ ЛЮДСЬКИЙ МОЗОК ЩОБ ЙОГО МОЖНА БУЛО Б ВІДТВОРИТИ У МАШИНІ, ЩО МИСЛИТЬ

1. Вступ

2006-й рік був до мене особливо неласкавим — із 13 статей, відправлених на міжнародні конференції цього року, 12 повернулися назад відхилені. І не просто так — “відхиленіми”! Всі, без винятку, — “Рішуче відхилені!”. Програмна комісія ECCV (Європейська Конференція з Комп’ютерного Бачення) була зі мною особливо суворою: “*Це філософська стаття... Проте, ні в правилах ECCV, ні в її традиціях немає місця для такого роду статей. Вельми шкодуємо (Sorry)*”.

Боже милосердний! Як же таке може бути в твоєму світі? Хоча, з іншого боку, неприязнь цих людей (моїх анонімних рецензентів і членів усіляких програмних комісій) можна легко і зрозуміти, — адже я, справді, у всіх своїх справах і писаннях прагну зайняти якомога більше відсторонену, філософську позицію.

Філософія для мене не лайливе слово. *Філософія для мене* — це природна потреба поглянути на речі якомога загальніше, щоб не втратити перспективи, ширшого погляду на предмет обговорення, і не впасти жертвою вузьких приватних визначень, пасток локального пошуку в обмеженому просторі або алхімії нескінченного перебору випадкових рішень, відповідних тільки для дуже обмеженого кола завдань.

Особливо цей філософський погляд важливий у таких, здавалося б, повсякденних, загальнодоступних і буденних речах, як зір. Ач дивина — людський зір! Будь-кому доступно.

Будь-хто ним володіє. Для будь-кого, здається, немає нічого простішого: подивився, побачив і — все зрозумів... Проте ж... Будь-які спроби передоручити цю справу машині (електронній відеокамері з приставленим до неї комп’ютером) ось уже півстоліття нікому не вдаються і повсюдно кінчаються повним провалом. (Навіть великі розумники з ECCV аніскільки в цьому до цих пір не досягли успіху, хоча традиційно пнулися безмірно. Ще б пак — Європейське Співтовариство витрачає сьогодні на ці речі величезні гроші, більше, ніж колись витрачали США, готуючись висадити людину на Місяць. Пам’ятаєте “Аполлон”?! Так зараз це у багато разів грандіозніше, особливо по грошах).

Традиційно вважається, що якість картини прямо пов’язана з роздільною здатністю відео-системи, — чим вона вища, тим краще ми бачимо. Чим більше деталей, тим вища якість зображення, тим воно (мабуть) цінніше для споживача. Успіхи сучасної електроніки спрямовані саме в цей бік: постійне, максимально можливе збільшення роздільної здатності фотоприймачів — від дешевих цифрових фотоапаратів з матрицею в 5-8 Мегапікселів (піксель — від англійського “*pixel*”, що означає “*picture element*”, окремий елемент картини) через професійні камери з розширенням до 16 Мегапікселів і до спеціальних відеокамер медичного, військового й аерокосмічного призначення з розширенням від 30 до 85 Мегапікселів. Зрозуміло, для відображення таких картинок потрібні відповідні екрани

спеціального (високого) розширення. І, звичайно ж, такі екрани вже є як для спеціальних застосувань (військових, медичних), так і для широкого вжитку: наприклад, стандарт HDTV (Телеба-



Еммануель (Амік) Діамант
науковий співробітник
ГАО НАН України (до 1971 р.),
старший науковий співробітник
Ізраїльського центру ядерних
досліджень (до 2000 р.),
м. Кир’ят Оно,
Ізраїль

emanl.245@gmail.com
<http://www.vidia-mant.info>

чення Високої Роздільної Здатності), обов'язковий для всіх телевізорів нового покоління, передбачає розширення в 1920x1080 пікселів (проти звичайного 720x576) або екрани для персональних комп'ютерів — 1280x1080 пікселів.

При цьому, звичайно, народна мудрість про те, що “бачать не очима, а розумом” (у російській мові етимологія слова “видеть” ведеться від слова “ведать”, — тобто “знати, розуміти”, а зовсім не від “виду” (що є “форма, розмір” або щось подібне), — все це, звичайно, в розрахунок не беруть. А шкода. Тому що на відміну від машини людина бачить не набір пікселів, а їх зв'язані, осмислені об'єднання, які прийнято називати “образами”, “об'єктами”, “предметами”.

Як і де ці набори окремих пікселів перетворюються на цілісні предмети, які ми осмислено сприймаємо і якими ми маніпулюємо в нашій свідомості? На це запитання ні у кого сьогодні немає відповіді. Тобто, на питання “де?” відповідь є: “у голові”, “у думці”. Але це мало чим комусь може допомогти. Тому що, що таке “розум” — теж поки ніхто не знає. А тоді запитання “як?” і пристосувати ні до чого.

2. “Ти ж мене підманула ...”

Зневажливе ставлення до філософських проблем зору миттєво обертається для нас вельми чутливими проблемами повсякденного використання машинного або, як його ще часто називають, комп'ютерного зору, який створений для того, щоб максимально задовольнити наші зорові запити і потреби.

Людина — істота, що постійно відчуває інформаційний голод. “*Infovore*” (що означає *інформаційно-голодний*) — так визначив це Ірвінг Бідерман, один із батьків-засновників науки про комп'ютерний зір (Biederman, 2006). Можливо, тому ми завжди так жадібно вдивляємося в телевізор. Може, саме тому таким успіхом користуються у нас відеофони і камерафони (camera-phone — відеокамера із мобільним телефоном). Одна тільки Nokia в 2007 р. продала 440 мільйонів таких телефонів з відеокамерами (Nokia, 2008). Це 40% світових продажів, тобто більше мільярда відеофонів було продано в одному тільки 2007 р. У 2009 р. їх буде продано ще більше — більше, ніж було виготовлено будь-якого типу фотокамер за всі роки з моменту винаходу фотографії (Thevenin et al., 2008).

Все разом це призвело до небаченого досі *потому відеоінформації*, який вирує навколо нас. Щоб якось спра-

витися з цим потоком, терміново вживають заходів зі стандартизації правил організації і регулювання цього потоку, тобто правил кодування й обміну візуальною інформацією, яка циркулює в цьому потоці.

Звісно, особливості людського зору слід тут взяти до уваги. Сумнівів із цього приводу ні у кого ніколи не виникало — комісія з розробки нового стандарту кодування відеозображень MPEG-4, розробка якого почалася в 1994 році, а остаточна редакція була завершена в 1999 році, урочисто проголосила: Новий стандарт буде об'єктно-орієнтованим (тобто кодуватися в ньому будуть не пікселі, а візуальні об'єкти). Для цієї мети в стандарт була введена ціла низка нових понять: VO (Visual Object), VOP (Visual Object Plane), VOL (Visual Object Level). (Puri & Eleftheriadis, 1998).

Ввести-то їх ввели, та тільки, як шукати їх і знаходити в картинці, — не сказали. Тому, що не знали. І дотепер не знають. А тому весь подальший розвиток і удосконалення стандарту (а за тих, що пройшли, 10 років, зроблено тут зовсім немало) пішов по шляху негласного удосконалення піксель-орієнтованого кодування. Об'єктно-орієнтований MPEG-4 насправді є модифікованим піксель-орієнтованим MPEG-2, хоча його сором'язливо і перейменували в MPEG-4/10 Advanced Video Coding Standard (H.264/AVC). По суті це є модифікований піксель-орієнтований MPEG-2. А у цього, зрозуміло, є свої невідворотні (проте, стандартом узаконені і до повсюдного застосування обов'язкові) сумні наслідки.

Попіксельна обробка картинок вимагає витрат часу й енергії. Особливо, якщо розміри картинки невинно ростуть. Це призвело до створення абсолютно нових типів оброблювальних пристроїв, т. з. цифрових процесорів, призначених для максимально швидкої переробки даних у картинках. Таким, наприклад, є Analog Devices TigerSHARC ADSP-TS201S процесор (TIGERSHARC — англійською Тигрова Акула), з продуктивністю в 3.6 GFLOPs (Гігафлоп — Гіга (одиниця з дев'ятьма нулями) обчислювальних (з плаваючою комою) операцій за секунду). Для великих картинок (HDTV) і цього, звичайно, мало. В цьому разі компанія BittWare пропонує PCI Mezzanine Card (спеціальну надбудовну плату), на якій розташовуються відразу чотири TigerSHARCа. І це ще не все — в одному комп'ютері можна встановити до чотирьох таких плат із загальною продуктивністю 57 GFLOPs (Bittware, 2007). Скільки ж при цьому

енергії витрачається? Краще не питайте. Хоча насправді, кого це цікавить?!.. Адже такі рішення призначені для стаціонарних пристроїв, де проблеми енергоспоживання (і охолодження, в свою чергу) мало кого хвилюють.

А як же бути з пересувними пристроями? — запитаете ви. — Як бути з нашими улюбленими мобільними відеофонами? Адже туди навіть одиниці TIGERSHARC не засунеш?..

Відповідь на це запитання виявляється надзвичайно простою. Тоді як роздільна здатність фотоприймача вашого відеофону безперервно росте (про що вам весь час нав'язливо нагадує реклама) — від 1.5 Мегапикселів (1280x1024) до 5 (2580x1930), 8 (3264x2444), 12 (4220x2820) і нарешті до 14 Мегапикселів (4570x3050) — сьогодні вже є і такі, — ефективний розмір картинки, яку ви можете спостерігати і передавати далі, міняється зовсім трохи: 80x60 пікселів, потім 160x120, і нарешті 352x288 пікселів (у США — 352 x240 пікселів) — це вже новітній Common Intermediate Format Standard (CIF). От так, поважані споживачі. Не хочете, можете не купувати, адже вас ніхто не примушує — ринок вільний. (Та ще який ринок! Сотні мільярдів доларів на рік! Програма “Аполлон” коштувала 25 мільярдів, щоправда, в цінах майже 40-річної давності.).

Сумна повість про те, як на цьому вільному ринку законно надувають вільних і суверенних у своєму виборі споживачів, на цьому не закінчується. Наступна історія стосується вибухового розповсюдження систем відеоспостереження з метою військової (оборонної), цивільної (суспільної) і, звичайно ж, особливої безпеки. Передбачено, що зростання обсягу продажів, пов'язаних з установкою систем відеоспостереження, виросте з 4.9 мільярдів доларів у 2006 році до понад 9 мільярдів доларів у 2011 році (Video surveillance, 2007).

Вважають, що рушійною силою цього вибуху є широке розповсюдження і впровадження мережевого Інтернет Протоколу ІП (IP — Internet Protocol) — своєрідної технології, що дозволяє зв'язати в одну-єдину систему розсіяні в просторі (або по території) відеокамери і центр (центри) їх комп'ютерного обслуговування. Одна тільки маленька непогодженість — ефективний розмір картинки, яка передбачена Інтернет-Протоколом, визначеним CIF Стандартом в 352x288 пікселів. Мало, дуже і дуже мало. Особливо, якщо врахувати, що картинка ці призначені для розгляду людиною. Бо сьогодні тільки людина, що

розглядає картинку, може зрозуміти, що ж там (на картинці) є або що ж там відбувається. А це означає, що для ефективної роботи системи відеоспостереження, до виходу кожної системи має бути завжди (й обов'язково) представлена людина — 24 години на добу, 7 днів на тиждень, 52 тижні на рік... Без цього нічого працювати не буде, без цього ніякої відеобезпеки ви не отримаєте. А людина — дуже тендітне створіння, більше п'яти хвилин на порожній екран (тобто, екран, на якому нічого не відбувається) дивитися не може, засинає з розплющеними владштована). Проте це мало кого хвилює — бізнес є бізнес, і попит задовольняється пропозицією. До того ж, щоб заощадити на робочій силі, черговому спостерігачеві на один екран, як правило, виводяться 4, 8 і навіть 16 камер (CIF навіть дуже в цих випадках доречний). Та й яка різниця? Ефективність нині чинних систем відеоспостереження у будь-якому випадку вельми сумнівна.

3. “Ой, казала мені мати та й наказувала...”

Створення відеосистем, здатних без допомоги або участі людини, самостійно розуміти, що вони бачать, і осмислено реагувати на побачене (хоч би на рівні привернення уваги оператора до ситуації, що представляє можливу загрозу) насправді завдання зовсім не нове і точно вже не мною придумане і поставлене. Початок цьому було ще покладено на Дортмундській зустрічі в 1956 році (McCarthy, et al. 1955), де четверо батьків-засновників (Маккарті, Мінскій, Рочестер і Шеннон) запропонували програму досліджень у галузі, яка з того часу називатиметься *Штучним Інтелектом* (Artificial Intelligence), і яка упродовж півстоліття займатиметься проблемами *Думаючих (Розумних) Машин* (Thinking Machines), як їх ще до цього задумав і визначив Алан Тюрінг, основоположник обчислювальної техніки і майбутніх комп'ютерів (Turing, 1950).

Як пов'язані між собою, запитаєте ви, системи відеоспостереження і автоматичного аналізу побаченого з Думаючими (Розумними) Машинами і Штучним Інтелектом? В тому і річ, що пов'язані, і навіть дуже. Про те, що “видеть” — це “ведать”, вже писалося. А ось про те, що половина людського мозку зайнята переробкою візуальної інформації, вам, звичайно, цікаво буде дізнатися (Milner & Goodale, 1998).

Академік *Ренін* вважає навіть, “що

мозок людини на 80 відсотків завантажений зоровою інформацією” [13]. В усякому разі, для батьків-засновників це виглядало абсолютно природним — вивчення проблем мозку потрібно починати з вивчення проблем зору. І жодних проблем у цьому ніхто з них тоді не бачив. *Стів Гранд*, наприклад, виступаючи на Panel Discussion at the Artificial General Intelligence Workshop у 2006 році, розповів таку історію: “У Родні Брукса (одного з корифеїв Штучного Інтелекту) зберігається меморандум, представлений *Мервіном Мінскім* (ще один із основоположників, ініціатор Дортмундської зустрічі), в якому він (Мінскій) пропонує (мова йде про кінець 1950-х) запустити проект, який вирішить усі вже виниклі на той час проблеми зору, найнявши для цього на час літніх канікул когось із старших студентів. “Я не знаю, — каже *Стів Гранд*, — де тепер цей студент. Але я думаю, що свою роботу він до цього часу так і не завершив” (Panel Discussion, 2006).

Зір, всупереч загальному переконанню, раптом виявився надзвичайно складною проблемою. Та і багато інших вихідних (початкових) передумов теорії Штучного Інтелекту раптом виявилися вельми сумнівними, чому прекрасна будівля науки про Думаючі Машини почала буквально на очах руйнуватися і сипатися, розпадаючись на дрібні, окремі субдисципліни. Так трапилося, наприклад, з Машинним Навчанням. Спочатку воно вважалося підрозділом Штучного Інтелекту, але дуже швидко перетворилося в самостійну дисципліну. При цьому точного визначення, що таке Машинне Навчання і як воно співвідноситься зі Штучним Інтелектом, до цього часу так і не існує. *Мартін Хьюттер* у своїй статті “Універсальний Інтелект: Означення Машинного Інтелекту” (Legg & Hutter, 2007) наводить 70 з гаком означень Машинного Інтелекту, взятих ним із праць авторитетів, що нині є здоровими й активно працюють у цій галузі. Жодної згоди між цими означеннями немає (і бути не може). Наявний різнобій можна пояснити тільки одним — відсутністю узагальнювального філософського підходу, наївною вірою в те, що з розрізненої дрібної цегли може бути зібрана велика будівля Машинного Розуму.

В умовах такого різнобою центр тяжіння наукових досліджень природно зміщувався, і змістившись, поступово поступився своїм місцем вивченню проблем так званого “біологічного зору”. Раптом стало модним навчатися у великого творця всього сушого в цьому світі — у Природи. До кінця

1970-х - початку 1980-х років оформилася й утвердилася у науці сучасна школа біологічного зору, яка на десятки років вперед визначила шляхи прогресу у всіх галузях, пов'язаних з вивченням роботи мозку. У тому числі і з розумінням функцій “комп'ютерного зору” або Інтелекту Думаючих (Розумних) Машин.

Провідними тут виявилися роботи *Давіда Марра* (Marq, 1978; Marq, 1982), *Анни Трейзман* (Treisman & Gelade, 1980) та *Ірвінга Бідермана* (Biederman, 1987). Немає сенсу заглиблюватися в деталі їхніх праць, але буквально двома словами слід було б викласти їхню основну ідею. Зводиться вона до наступного: *Зір* (за Анною Трейзман) є взаємодією двох потоків обробки інформації. Один із них “висхідний”, спрямований “знизу-вгору” (bottom-up directed) процес (від очей, через “нижній”, найближчі до очей ділянки мозку, до “верхніх”, свідомих, ділянок, які думають), в якому несвідомо виділяються та обробляються дрібні (піксельні) елементи інформації, котрі об'єднуються в більші проміжні елементи картинки (image features). Другий потік — “спадний”, спрямований “згори-донизу” (top-down directed) процес, в якому ці початково виділені елементи тепер уже осмислено об'єднуються (групується) в цілісні об'єкти (предмети). Саме ці предмети ми і “бачимо”, саме ними ми і оперуємо, коли “розглядаємо” картинку, намагаючись зрозуміти, що ж там зображено (Treisman & Gelade, 1980).

Принципи роботи висхідного потоку з самого початку були легко та інтуїтивно зрозумілі кожному, досить просто формалізувалися математично, а тому швидко й охоче були засвоєні та підхоплені всіма, серед інших і розробниками “машинного” або “комп'ютерного” зору. З того часу тільки цим і займаються, намагаючись усі проблеми машинного зору звести до рішення цих “нижньорівневих” завдань (low-level image processing tasks, low-level bottom-up pixel-oriented image processing).

Принципи роботи спадного потоку з самого початку були вельми туманними та розмитими. Зрозуміло, що для успішного об'єднання розрізаних деталей в якесь загальне ціле потрібне якесь особливе початкове знання принципів, за якими це об'єднання мало би проводитися. Зрозуміло, що розташовується це знання в “найвищій”, “свідомій” зоні мозку (і тому весь процес вважається спрямованим “згори-вниз”, тому він називається “спадним”).

Гостра необхідність зрозуміти принципи, за якими має здійснювати-

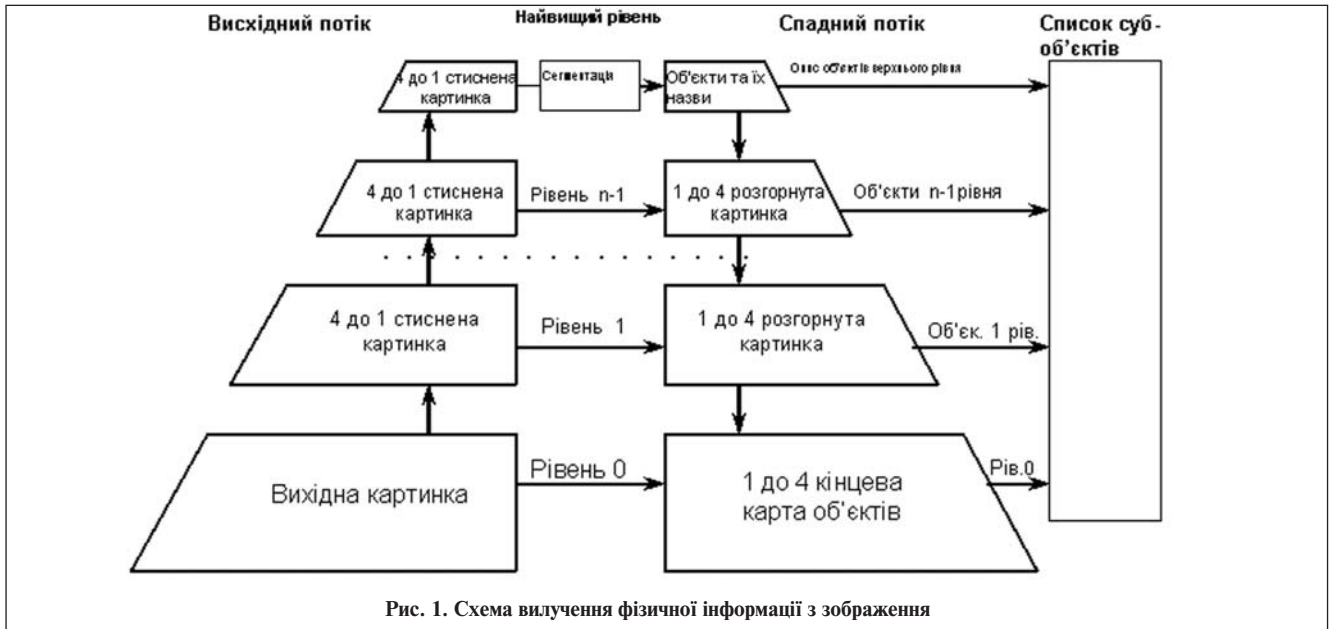


Рис. 1. Схема вилучення фізичної інформації з зображення

ся це об'єднання, призвела до виникнення та інтенсивної розробки цілої нової галузі знань, що отримала назву *Проблеми Зв'язування (The Binding Problem)*. Її рішення було присвячено (і сьогодні присвячується) маса засобів і зусиль. Багаторічним марним працям на цьому терені присвячений спеціальний випуск журналу *Neuron* (Volume 24) за 1999 рік, а також спроби самої *Анни Трейзман* якось просунути в цьому напрямку (Treisman, 1996). Коли виявилось, що всі ці зусилля ні до чого не ведуть, *Кох* і *Крік* (так, так, той самий Крік, Нобелівський лауреат, який разом з *Уотсоном* відкрив подвійну спіраль ДНК) запропонували *ідею "гомункулуса"*, маленької людини, яка сидить у нас в голові, і все знає, все за нас вирішує. Зокрема, і як розумно зв'язувати розрізнені елементи картинки в цілісні об'єкти і предмети (Crick & Koch, 2000).

Теорія гомункулуса не прижилася. Але потреба в "зв'язуванні" залишалася і навіть ставала все гострішою - без вирішення проблеми зв'язування неможливо було виконати об'єктно-орієнтований опис картинок. А без цього не можна вирішити головну проблему сьогоdnішнього Інтернету — проблему "людиноподібного" пошуку в мережі потрібного матеріалу за його візуальним змістом. У спеціальній літературі це називається *Пошук Зображень за змістом (CBIR — Content-Based Image Retrieval)*. Оскільки навіть теорії біологічного зору не можуть пояснити, як з окремих елементів картинки формуються осмислені візуальні об'єкти, а без цього ні про який осмислений (семантичний) опис змісту картинки не може бути і мови, а отже, пошук за змістом (за сенсом)

стає безглуздим. Тоді — навіщо взагалі нам всі ці теорії?! Тоді — Даєш алхімію! Даєш єдино доступне і таке, що очевидно напрошується, рішення — з'ясної і доступної інформації нижнього рівня здобуватимемо недоступну нам інформацію верхнього рівня! (Щоб ви не подумали, що я знову в черговий раз роблю наклеп на добрих людей, пошлюся лише на декілька праць у цій галузі, взятих мною на вмання: Mojsilovic & Rogowitz, 2001; Zhang & Chen, 2003; Itti, 2005; Serre et al., 2005; Nare et al., 2006; Kveraga et al., 2007. Насправді ж цей список нескінченний.)

Дивлячись на все це свято життя, я тільки і можу сказати: "Мені дуже шкода, але людський зір і людське мислення влаштовані зовсім не так і зовсім не так працюють..."

4. "Путь далёк у нас с тобою..."

Я не хочу сказати, що я завжди був таким "розумним", як це може кому-небудь здатися сьогодні. У жодному разі. Свої перші кроки у вивченні машинного зору я починав, як і всі, окрилений ідеями *Марра* про "Початковий" або "Двох-з-половиною-вимірний Ескіз", які (згідно з теорією *Марра*) найкоротшим шляхом мали привести нас до якнайкращого опису змісту картинки. (Під змістом картинки малася на увазі, звичайно, інформація, що міститься в картинці).

"Бачите, — азартно агітував я своїх керівників, прагнучи переконати їх виділити мені якісь бюджетні кошти і час на проведення потрібних (у моему розумінні) досліджень. — Ви бачите, як декількома штрихами, декількома

скупими лініями художник передає вам повне уявлення про предмет, який він зобразив на картинці. Контурні і прикордонні лінії є основними носіями інформації в картинці. Навчившись виділяти і обробляти їх, ми отримаємо ключ до розуміння картинки!"

Мої керівники скептично посміхалися і не поспішали ділитися зі мною своїми бюджетами. Проте, дещо вони все-таки мені дозволяли робити. І все для того, щоб я міг сам переконатися, що моя віра і мій оптимізм дуже перебільшені, — замість ключів до розуміння картинки я навчився отримувати купу краєвих елементів, що виділяються в результаті прогону оператора 3x3 (або 5x5, або 7x7) по полю картинки в 256x256 пікселів. (У режимі майже реального часу.) Всі сили і час йшли тільки на це. А в результаті? Пшик, безформна купа краєвих елементів (edges), які неможливо ні сортувати, ні яким-небудь чином групувати або зв'язувати (пам'ятаєте "The binding problem"?), щоб отримати хоч якусь подібність контуру окремого об'єкта або предмета.

Це був жах, і не тільки мій особистий жах. Багато людей до цього часу борсаються в цьому болоті, не маючи сил звільнитися від нав'язливих ідей про цінність прикордонних ліній. Багато хто до цього часу випробовує долю, намагаючись (сподіваючись) знайти вихід зі становища. В усякому разі, потік публікацій і переможних реляцій на цю тему (знайшов! знайшов!) не закінчується до цього дня (Ghosh et al., 2007; Awad & Man, 2008; Qiu & Sun, 2009).

Проте двадцять років тому загальна картина виглядала зовсім не так похмуро. Хоч я разом зі всіма і метушив-

ся на “висхідному”, що йде “знизу-вгору” магістральному шляху піксельної обробки картинок, думка про заповітну, що йде “згори-вниз”, піксель-зв'язувальну ідею не покидала мене ніколи. Не дуже чітко уявляючи собі, що це означає, я все-таки вчився і намагався витягувати з картинки Інформацію.

А що це таке — “інформація”?

Що таке “Візуальна” або “Зорова” інформація?, — як і всі мої співтовариші (тоді і зараз), я не дуже точно уявляв собі, що це таке. Мої вчителі — *Марр*, наприклад, який першим увів у вживання термін “зорова інформація” (*Marq, 1978*), теж не дуже піклувалися про те, щоб точно визначити, про що ж, власне кажучи, йде мова. Нікому, щоправда, це ніколи не заважало. І мені теж. Я навіть встиг у ці роки винайти для себе таке поняття, як “Інформаційний Зміст Окремого Пікселя” (*Single Pixel Information Content*) і міру його кількісного визначення (*Diamant, 2003*). Експериментуючи з цим заходом, я винайшов і навчився обчислювати “Питомий Інформаційний Зміст Картинки” (*Image Specific Information Density*) і несподівано для себе виявив “Принцип Збереження Інформаційного Змісту Картинки” (*Image Information Content Conservation Principle*). Згідно з цим принципом, при зменшенні розміру картинки, її питомий інформаційний зміст не міняється (до певної межі), а в деяких випадках навіть трохи росте. Після досягнення якоїсь межі стиснення, її питомий інформаційний зміст різко падає. Ось цей передкритичний розмір картинки, вважав я, і має бути тим оптимальним розміром, в якому треба здійснювати пошук інформації в картинці. (Щоб не витратити даремно сили і часу на перебирання всіх пікселів у картинці. Пам'ятаєте, звичайно, *TigerSHARC?*)

Своє чудове відкриття я, зрозуміло, спробував негайно обнародувати, але, як завжди, був “рішуче знехтуваний”. Одну з таких спроб — подачу на *BMCV-2002* (*British Machine Vision Conference, Британська Конференція з Машинного Зору*) — я дбайливо зберігаю на своєму сайті, і посилання на нього (*Diamant, 2002*) наведено в списку моїх джерел. Найцікавіше у всій цій історії, що таку роботу було зроблено і такі результати були отримані потім співробітниками *MIT* — *Массачусетського Технологічного Інституту* [1]. Щоправда, на сім років пізніше. І в результаті психофізичних експериментів над людьми, а не кількісних (як у мене) вимірювань.

Проте — “не для цього доводи”.

Головне, що мені вдалося відірватися від панівних традиційних уявлень і продовжити розвивати ту позицію, яка здавалася мені доречнішою. До цього часу було вже зрозуміло, що пошук інформації в картинці потрібно починати з її дуже стислого початкового уявлення. До цього ж підштовхували і відомі експериментальні дані, отримані в біологічних дослідженнях. Щоправда, якщо не сліпо слідувати за ними, а намагатися давати спостережуваним фактам свої пояснення.

Річ у тому, що технологи машинного зору зовсім не дарма погналися за високою роздільною здатністю своїх фотоприймачів. Від біологів їм уже давно відомо, що розгляд спостережуваної сцени людина здійснює, послідовно скануючи її поглядом. При цьому в кожен такий момент на потрібний об'єкт наводиться центральна ділянка ока, так звана “fovea”. (На відміну від фото- або телекамери, фотоприймачі (пікселі) людського ока розподілені по сітківці ока дуже нерівномірно. Дуже висока щільність фотоприймачів припадає на дуже маленьку центральну ділянку сітківки, яка і називається “fovea”.) Ось її-то, “fovea”, і спрямовує око (мозок) на розгляд потрібного предмета. Через неї і надходить до нас головна частина зорової інформації. З цієї причини всі дослідження в галузі біологічного зору стосуються винятково досліджень “fovea-vision”, а всі творці електронних фотоприймачів намагаються тільки його й імітувати.

Проте при всьому цьому одне питання абсолютно випадає з обговорення: якщо вся інформація надходить до нас через “fovea”, яким чином око (мозок) знає, куди саме спрямувати “fovea” в кожен такий момент? Адаже сканування сцени здійснюється не за раз і назавжди заданим циклом, не за растром (як на телебаченні, скажімо), а якось дуже навіть мимоволі, безсистемно. Звідки ж мозок знає, куди саме мають дивитися очі в кожен такий момент? Біологи не дають відповіді на це запитання. (Втім, вони його собі і не задають.)

Але можлива відповідь на нього могла б бути: в мозку у того, хто дивиться, напевно, є загальна карта “місцевості”, по якій він вирішує, яку саме її ділянку йому слід було б розглянути детальніше (куди спрямувати свою “fovea”). Розглядаючи і вивчаючи сцену, людина одночасно користується двома картами: великого і малого масштабу. “Периферійні” ділянки сітківки забезпечують його менш докладними картами, “fovea”- докладнішими.

Між іншим, телевізійний оператор, камера якого має тільки фотоприймач максимально можливої роздільної здатності, розв'язує це (можливо, класичне) загальне для всіх завдання за допомогою трансформатора — спочатку знімається найзагальніший план, а потім камера “найджжає” на потрібну частину сцени і зосереджується на її деталях.

Весь цей довгий реверанс у бік біологічного зору потрібний мені був тільки для того, щоб ще раз підкреслити те, що вже напрошується само собою: *пошук інформації в картинці має починатися із найзагальнішої “карти місцевості”, з найбільше стислої та зменшеної копії картинки.*

Це, звичайно ж, суперечить всім відомим теоріям зору, за якими пошук інформації в картинці має проводитися знизу-вгору, від піксельних деталей до смислового загального.

Оскільки одними тільки емпіричними висновками тут вже не можна обійтися, потрібно було терміново шукати теорію, яка могла б підтримати і зміцнити цей напрям думки.

Дуже швидко виявилось, що така теорія є. І навіть не одна, а цілих три — десь у середині 1960-х років, приблизно одночасно, але абсолютно незалежно один від одного, були зроблені й опубліковані три роботи, які спочатку і не привернули до себе особливої уваги: “Формальна теорія індуктивного виведення” *Р. Соломоноффа* (*Solomonoff, 1964*) “Три підходи до чисельного визначення інформації” *А. Колмогорова* (*Kolmogorov, 1965*) і “Про довжину програм для розрахунку кінцевих бінарних послідовностей” *Г. Хайтіна* (*Chaitin, 1966*). Оскільки зі всіх трьох — Колмогоровська стаття найвідоміша і найпопулярніша тепер, я в своїх подальших міркуваннях посилатимуся тільки на неї.

Як і Теорія Інформації *Шеннона* (*Shannon, 1948*), опублікована майже за двадцять років до цього, Колмогоровська теорія була спрямована на пошук шляхів *вимірювання кількості “інформації”*. Проте тоді як Шенноновська теорія була зайнята оцінкою середньої кількості інформації, отриманої на виході джерела з випадковим розподілом сигналу, Колмогоровська теорія була зосереджена на інформації, що міститься в одному окремому ізолюваному об'єкті. На мій погляд, це значно більше підходило для обговорення проблем, пов'язаних з особливостями людського зору.

Як і у випадку з *Марром* і *Анною Трейзман*, я не стану навантажувати моїх читачів докладним викладом Колмогоровської Теорії Складності.

Про це ви можете і самі прочитати сьогодні в чудових текстах *Пауля Вітані*, а також у працях інших численних учнів і послідовників *А. Колмогорова* (Li & Vitanyi, 2008; Grunvald & Vitanyi, 2008). Мої наміри набагато скромніші: спираючись на ідеї Колмогоровської Складності, спробувати поглянути по-новому на проблеми людського зору і пошуку інформації в картинках. Наполегливо повторюючи при цьому: Колмогоровська теорія — чиста математична теорія, в якій відсутня біологія. В наші дні, коли будь-яка ідея в галузі машинного зору або штучного інтелекту поспішає оголосити себе “Biologically inspired”, я повторюю знову і знову: “Шукати пояснення потрібно в чистій логіці, в математиці! Ніякі “біологією навіяні” пояснення нам не допоможуть!..”

Як це все відбулося на моїх пошуках інформації в картинках і до чого все це мене врешті-решт привело, ви дізнаєтеся, якщо у вас вистачить терпіння дочитати до кінця цю статтю.

5. “Точка, точка, запятая — вышла рожица кривая”

Отже, перша і найголовніша новина, яку я почерпнув для себе з теорії Колмогорова, була: *інформація — це опис*. Знаковий (буквений, цифровий) або складніший лінгвістичний (мовний) опис, за яким предмет опису може бути достатньо точно відновлений і

відтворений. *Колмогоров*, а особливо його учні, прирівнюють такий опис до комп’ютерної програми, яка керує реконструкцією початкового предмету опису (Vitanyi, 2006). У картинці такими предметами, безперечно, є структури, утворені конгломератами пікселів. Колмогоровська теорія вказує, як саме треба створювати описи таких структур. Спочатку створюються узагальнені, спрощені описи. На наступному етапі ці вихідні дані починають обростати подробицями, стають більш деталізованими. Цей процес повторюється на всіх подальших, нижчих рівнях, таким чином, теоретично, весь процес може йти нескінченно. Проте стосовно зору він іде тільки до того рівня, де ступінь деталізації забезпечує системі ухвалення найоптимальнішого рішення.

Цей підхід дуже нагадує інший давно відомий принцип *Оккамової бритви* (*Occam's Razor*): “*Зі всіх можливих припущень, що описують конкретне спостереження, завжди вибирай найпростіше*” (Sadrzadeh, 2008).

Таким чином, слідуючи Колмогоровській теорії, ми можемо заявити, що інформаційний опис, насправді, *це не якийсь єдиний монолітний опис, а ціла ієрархія описів, де подробиці та деталі опису множаться на кожному нижчому рівні ієрархії*. Іншими словами: починаючи з найбільш узагальненого (спрощеного, стилізованого) опису, ієрархія інформаційного опису розгортається “згори-вниз”, відповідно

до принципу “від-загального-до-деталей”. (Увага! Увага, панове! Жодних потоків, що йдуть “знизу-вгору”, тут не згадано, і ні про які “проблеми зв’язування” тут немає і мови! Все зовсім навпаки! Ви розумієте, що це означає?!). А це означає, що реальний процес витягання інформації, що йде “згори-вниз”, з картинки, обходиться взагалі без будь-яких попередніх знань верхнього рівня! Тому, що знання верхнього рівня жодного відношення до витягання інформації з картинки не мають! Тому, що її (інформації верхнього рівня) там просто немає! Вона є і працює в голові у спостерігача, який розглядає картинку, але в самій картинці її немає.

Зрозумівши це, я зрозумів, що насправді *ми завжди маємо справу з двома видами інформації: об’єктивною (фізичною) і суб’єктивною (семантичною)*. Фізична інформація міститься в картинці та є описом структур, що виникають у масиві фізичних даних, які представляють цей об’єкт. У нашому випадку це пікселі, в інших випадках це може бути все, що завгодно. Важливо тільки, що це спостережувані структури реальних (фізичних) даних, які присутні в цьому об’єкті. І тому *опис цих структур є фізичною інформацією, яка належить об’єкту*. Інший вид інформації, з яким нам доводиться мати справу, — це опис взаємин і зв’язків, які є або можуть мати місце між окремими елементами фізичних структур, спостережуваних у

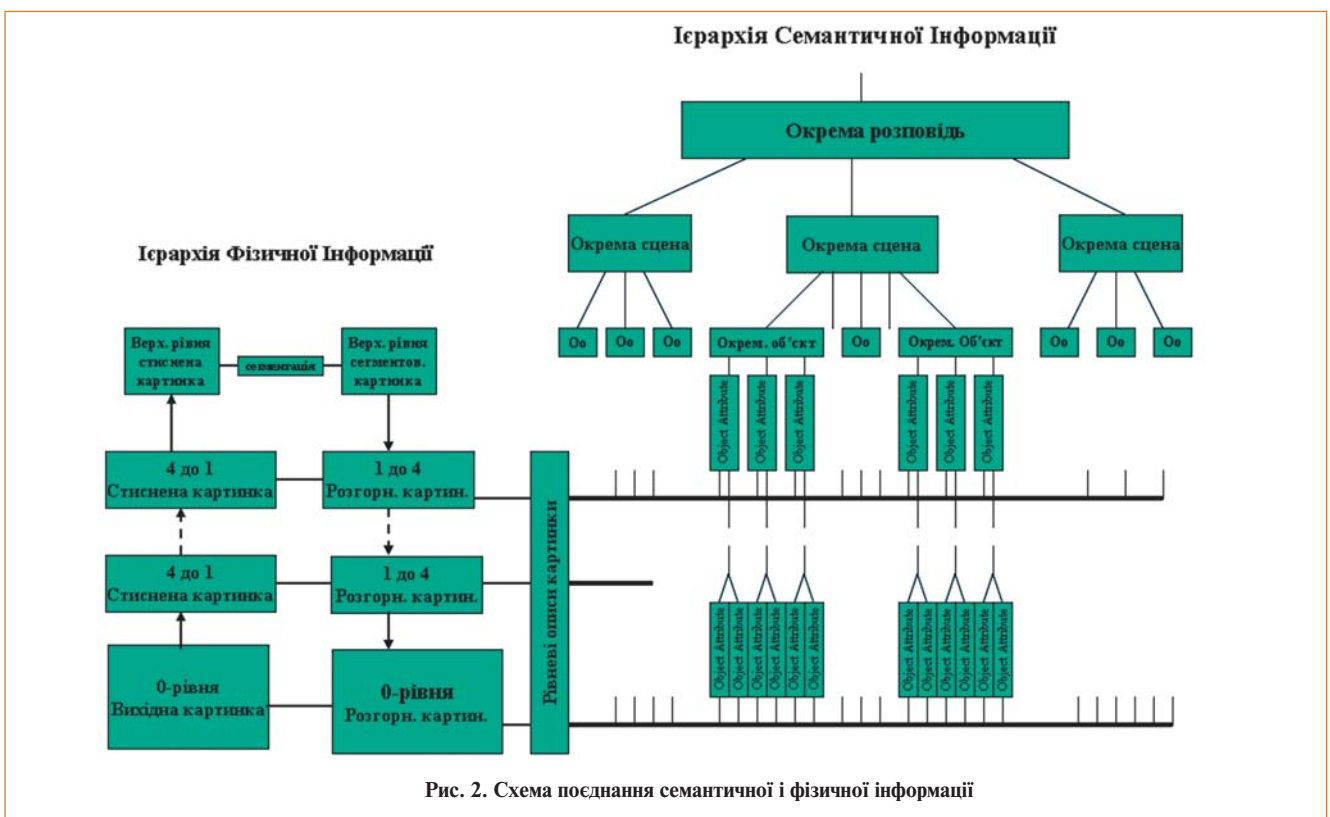


Рис. 2. Схема поєднання семантичної і фізичної інформації

цьому об'єкті. Це завжди наше (людське) пояснення цих зв'язків. Тому одна і та сама картинка (одна і та сама фізична інформація) може по-різному розумітись (тлумачитись) різними людьми. Тобто, одній і тій самій фізичній інформації може бути привласнена різна семантика, різна семантична інформація, тому що ця інформація належить не самому об'єкту, а тому суб'єктові (глядачеві, спостерігачеві), який його (цей об'єкт) розглядає. Тобто, *семантична інформація не міститься в картинці, а привласнюється їй спостерігачем* (у його, спостерігача, голові).

Це дуже важливе і чудове спостереження. Воно, звичайно ж, суперечить всім традиційним уявленням про інформаційний зміст картинки і способи його витягання. Але воно не суперечить логіці та здоровому глузду. Тому потрібно було терміново перевірити, як це виглядає на практиці. Як можна, слідує запропонованим мною принципам, провести сегментацію заданої картинки і отримати кінцевий набір (ні, не готових фінальних об'єктів, — адже це робота для людського мозку), а “зв'язаних” структурних субчастих цих (майбутніх) об'єктів. Я виконав усю цю роботу, і мені навіть вдалося опублікувати деякі результати моїх експериментів (Diamant, 2004; Diamant, 2005a; Diamant, 2005b).

Я вже сказав раніше (повторюючи це услід за Колмогоровим), що інформаційні описи повинні бути завжди реалізовані з допомогою якої-небудь мови опису. Хоча сьогодні існує вже дуже багато формальних мов опису, мені здається, що для конструювання думуючих (розумних) машин нам найкраще було б скористатися вже наявною мовою людського спілкування. Адже саме цією мовою користуються єдино знайомі нам і добре зарекомендовані розумні системи (системи людського розуму).

У цьому разі семантична ієрархія має виглядати таким чином. Скажімо, опис якого-небудь акту в п'єсі або сценарії завжди розпадається на описи його окремих сцен і мізансцен, а ті, у свою чергу, на описи окремих акторів і предметів, що беруть участь у кожній мізансцені. Деталізація описів на цьому не закінчується, і кожен окремий предмет, у свою чергу, може бути розкладений на складові компоненти.

Ось тут і вступає в гру фізична інформація, тому що деталі найнижчого рівня семантичної ієрархії це і є та фізична інформація, яку ми сприймаємо нашими органами чуття. (Ця фізична інформація, як ви, звичайно,

пам'ятає, сама по собі є ієрархією, що розгортається згори-вниз).

Узагальнюючи сказане, можна тепер упевнено заявити, що розуміння фізичної інформації, яку дають нам наші органи чуття (зір, наприклад), є вбудовування (прибудовування) цієї фізичної інформації у відповідну семантичну структуру, де така (або дуже близька до неї) фізична інформація вже присутня. Ми шукаємо в своїй пам'яті “розповідь” про предмет або подію, фізична інформація якої найкращим чином відповідає тій фізичній інформації, яку ми в цей момент спостерігаємо. Ця витягнута з пам'яті розповідь і є інтерпретацією спостережуваного в цю хвилину. І якщо такої відповідної розповіді в нашій пам'яті не виявляється, — ми не розуміємо, що ж ми бачимо, ми насправді просто нічого не бачимо. У біології (психології) зору це відоме явище, яке (перекладаю, як умію, з англійської) називається “*Cumyauціна Clinoma*” (*Situation Blindness*).

Цей пошук інтерпретації розповіді в пам'яті людського спостерігача може бути представлений алгоритмічно, тобто може бути реалізований машиною (комп'ютером). А це означає, що принцип роботи думуючих (розумних) машин нам тепер зрозумілий, і ми можемо негайно приступити до їх реалізації. Залишилося лише декілька дрібних деталей, які потрібно буде обговорити і вирішити вже по ходу справи... Проте, як мовиться, “*диявол-то, він завжди в подробицях*”.

6. “Что-то с памятью моей стало...”

Викладену щойно принципову схему думуючої (розумної) машини я опублікував уже кілька років тому (Diamant, 2007; Diamant, 2008), але не було жодної реакції на це. Тому продовжу і далі в гордій самоті...

Зі всього сказаного вище однозначно впливає, що основним елементом думуючої (розумної) машини (і розумної людини, звичайно) є пам'ять, в якій накопичуються розповіді про різні випадки в нашому житті. Кожну нову фізичну інформацію, яку нададуть нам наші органи чуття, ми негайно (і часто несвідомо) “прибудовуємо” в одну з тих розповідей, що вже є в нашій пам'яті, і якщо ця “прибудова” виявляється вдалою, ми негайно розуміємо (дізнаємося) з тексту розповіді, як же нам належить реагувати, тобто, що ж нам робити далі, бо в тексті розповіді вже міститься опис всього того, “що було, що буде, і чим серце заспоко-

їться”. А тоді виникає законне запитання: як же занести нам у пам'ять нашої машини всі ці (на всі випадки життя відповідні) розповіді? З людиною, здається, все зрозуміло — особистий життєвий досвід і розповіді досвідчених людей споконвіку становили істотну частину людської освіти, тобто заповнення її пам'яті потрібними розповідями. Причому ці розповіді зовсім не мали бути пов'язані безпосередньо з особистим досвідом конкретного індивіда. А що ж нам робити з машиною? Як її наділити людським досвідом? Відповідь на це дуже проста — штучно перенести (внести) в пам'ять машини ті розповіді, які будуть потрібні їй для розв'язання того кола завдань, які ми хотіли б їй передати.

Це, звичайно, дуже знижує наші шанси створити людиноподібну розумну машину, але в цих своїх мріях ми дуже часто і дуже сильно перегибаємо. Хіба люди всі однакові? Хіба інформація чи розповіді в пам'яті хірурга схожі на розповіді в пам'яті педіатра, або механіка в авторемонтній майстерні, або шеф-кухаря в ресторані? Хіба ці розповіді були отримані ними з особистого життєвого досвіду? Звичайно, ні! Вони (люди) всьому цьому навчені. Ці знання (ця інформація, ці розповіді) були ними отримані вже в готовому вигляді, ззовні. То чому ж до машини ми маємо ставитися інакше? Чому від машини ми чекаємо універсальності і “самонавчання”, які не існують навіть у людей?

Кажучи це, я наживаю собі нових ворогів, які сповідають і проповідують сьогодні теорії про самонавчання живих організмів у природі й інтелігентних машин у нашому розпорядженні. Ні, — стверджую я. — Ніякого самонавчання немає! Ні в природі, ні в штучно створюваних нами машинах. (Тобто іноді це трапляється, але не як правило, а як рідкісний виняток. Як *Ньютон* або *Ейнштейн*, наприклад, або які-небудь геніальні поети. Всі інші отримують свої знання завжди ззовні, завжди вже в готовому вигляді). Це добре видно в новітніх дослідженнях навчання, тобто, передачі інформації (знань) від однієї живої істоти до іншої (як серед бактерій так і серед вищих тварин) і цілеспрямованого навчання дитинчат у світі тварин. (Я докладніше говорю про це в своїх інших статтях, де цитую відповідні роботи на цю тему). Це тільки у *Кінлінга* Мауглі ходить на двох ногах і розмовляє як людина, — в реальних же умовах діти, що виростили серед вовків, наприклад, могли пересуватися тільки рачки і могли тільки вищати. Люд-

ській поведінці їх не навчили. Тоді як діти однієї раси, вирощені в іншому расовому середовищі, легко засвоюють усі правила поведінки і мову цього середовища. Тому потрібно швидше позбутися століттями освячених забобонів і швидше пересаджувати машинам ту частину людської пам'яті, ті конкретні розповіді, необхідні їм для вирішення певного кола завдань, які ми у кожному конкретному випадку хотіли б їм передоручити (наприклад, невтомно дивитися в порожній екран систем відео-розвідки і відео-спостереження).

Очевидною новиною в цьому процесі перенесення знань є те, що формою запису минулого знання я вважаю розповідь. (Треба було б ще до всього визначити, що таке “знання”. Не вдаючись до детального обговорення цього вельми делікатного питання, скажу тільки: для себе я визначаю “знання” як “зафіксовану (записану) інформацію”. Прошу саме це мати на увазі, коли вам трапляється термін “знання”).

Дуже цікавий аргумент на користь такого “літературного” підходу я знайшов нещодавно в роботі Ізраїля Мойсевича Гельфанда і його співавторів, присвяченій досвіду співпраці між лікарями і математиками в справі створення “розумної” машини для автоматичного діагностування захворювань [15]. Створення такої машини потребувало особливої мови спілкування між машиною і людьми. Зазвичай, такі мови створюються у вигляді онтологічного словника термінів, однаково зрозумілих і машині, і людям, які з нею працюють. (Основоположна стаття на цю тему опублікована Томасом Грубером тільки в 1993 році (Gruber, 1993). Гельфанд і його співробітники зіткнулися з цим і думали про це ще в 1980-х. До речі, Гельфанд — учень А.М.Колмогорова. Колмогоров прийняв його до себе в аспірантуру, хоча у Гельфанда не було навіть відповідної формальної освіти. Зате потім уже Академія наук СРСР “відігралася” на ньому за все — його, почесного члена багатьох іноземних академій, радянська Академія наук відмовлялася визнати, і за кордон представляти радянську науку його, звичайно, не випускали).

Так от що писав І.М. Гельфанд з приводу мови думаючих (розумних) машин: “...є два варіанти розвитку (літературної) мови: написання високохудожніх творів і складання тлумачного словника. Ми знаємо, який великий вплив на розвиток російської мови мали і Пушкін, і Даль” (Шекспір і Доктор Джонсон в англійській мові, додамо ми від себе).

У ті далекі 1980-і роки творці діагностувальної машини вибрали для себе шлях створення словника. Проте це не відмінило і не відміняє справедливості першої частини визначення, даного І.М.Гельфандом. Сенс його прозріння став мені зрозумілий лише після того, як я сам “дійшов” до розповіді як форми запису і представлення семантичної інформації (Ось вам ще один приклад “ситуаційної сліпоті”). Решта світу до цих пір зайнята складанням, зведенням воедино і удосконаленням онтологічних словників.

Кажучи про “літературу” (про “розповідь”) неможливо не задуматися про мову, якою все це має бути записано. Те, що інформаційний опис обов'язково має бути реалізований за допомоги якої-небудь мови, придумано зовсім не мною, — про це сказано ще у Колмогорова. Нам же слід тільки уважно простежити, куди нас можуть завести всі ці розмови.

А ведуть вони нас ось куди: якщо правильним є те, про що ми щойно говорили, то *пам'ять, в якій зберігаються “розповіді” будь-якої живої істоти, має бути функціонально дуже схожа на пам'ять комп'ютера, тобто бути такою, що в неї зручно вписуються (і зчитуються з неї) стрінги відповідних текстів. Є така пам'ять у живих істот? Звичайно, так! І ми з нею добре знайомі — це геном, наша генетична пам'ять.*

Хоча, якщо порівнювати з комп'ютером, то в останнього є два види пам'яті: *постійна й оперативна (hardware and software)*. Якщо це правило не випадковий артефакт, а справді має загальний характер, тоді й у живих істот теж мають бути такі ж два види пам'яті: постійна й оперативна. *Постійна ROM — Read Only Memory (пам'ять тільки для читання) і оперативна RAM — Random Access Memory (пам'ять випадкового звернення)*. Так воно і є! У кожній живої істоти (зокрема й у людини), є така *Постійна пам'ять* (тільки для читання), яка вже давно всім відома і знайома, — це наш геном, наша успадоквана генетична пам'ять. У людини ж є (і кожен може підтвердити це на підставі свого власного досвіду) й *Оперативна пам'ять* — те, що записується і зберігається у нас в голові протягом нашого індивідуального життя!

Я бачу, як грізно супляться мої опоненти, — адже інформація у нас в голові — це зовсім не стрінги, а пачки електричних імпульсів (так нас завжди учили нейробіологи), і зберігається вона зовсім не у вигляді стрінгів, а... Втім, як зберігається у нас наша пам'ять, ніхто не знає. Є, кажуть, у де-

кого якісь міркування, але все це дуже і дуже туманно.

Те, що будь-яка інформація, виходячи зі всього раніше сказаного в цьому тексті, є текстовим описом (а значить, стрінгами), сперечатися далі вже безглуздо. Що генетична пам'ять — це набір хімічно записаних текстів (тобто стрінгів), уже давним-давно не новина. Все це тільки ще раз підтверджує вищесказане. Те, що в оперативній пам'яті ми, швидше за все, маємо справу з аналогічним явищем, тобто інформація між нейронами записується і передається не як пачки імпульсів, а як хімічно записані фрази, — сьогодні теж уже відомо багатьом: астроцитами, наприклад, обмінюються інформацією з нервовими клітинами без будь-яких імпульсів (Volterra & Meldolesi, 2005; Haydon & Carmignoto, 2006). Саме тому їх до недавнього часу впритул не бачили — адже шукали, як завжди, Індію, і не хотіли помічати розташовану перед ними Америку. “Ситуаційна сліпота”, звісно.

“Але ж електричні імпульси між нейронами — це реально спостережуваний і вимірюваний факт?” — скажете ви. Правильно, і спостережуваний, і вимірюваний. Але ось вам інший добре відомий приклад: під час роботи струминного принтера з дюз його друкувочої головки “випльовуються” мікродози тонера (спеціального чорнила для принтера), які власне й утворюють зображення, що залишаються на папері. Ви можете ухитритися і заміряти цю імпульсну витрату чорнил, які принтер витрачає на друк заданого тексту. А зараз вирішите: чи зможете ви коли-небудь витягнути з цих вимірювань (імпульсної потужності чорнильних струменів) сенс того, що ви тільки що віддрукували на папері? Навряд чи. Пачки імпульсів, що розповсюджуються між нейронами, мають таке саме відношення до інформації, як пачки імпульсів тонера до надрукованого тексту.

До таких прикладів можна віднести і спостереження за тимчасовими змінами потоку електронів у електронно-променевої трубі осцилографа або телевізора. Добре було б, якби ці аргументи якось відклалися в голові моїх читачів.

7. “Хотел бы в единое слово...”

Зважаючи на вищесказане, мої відносини з професійними біологами (втім, як і з моїми колегами по комп'ютерному зору) залишаються вельми напруженими. А шкода. Ми могли б бути один одному корисні. Я маю на увазі не тільки нову осмислену інтер-

претацію тієї фізичної інформації, яку біологи сьогодні отримують у величезних обсягах, але не можуть її правильно тлумачити, тому що плутають фізичну і семантичну інформацію, не розуміючи (а тому не бачачи) різниці між ними. Є ще багато інших, більш загальних, філософських аспектів всього вищесказаного, які нам добре було б обговорювати спільно.

При цьому я зовсім не збираюся вдавати, що я не розумію, як далеко ми віддалені один від одного, що проголошений вище принцип (“згори-вниз”) інформаційного опису (“від загального до деталей”) прямо суперечить традиційному принципу (“знизу-вгору”) збирання окремих приватних свідчень, перш ніж хто-небудь зважиться робити з них відповідні висновки й узагальнення. Я розумію, що для багатьох це розуміння, яке йде “згори-вниз”, нагадує Божественний дух і провидіння, але запевняю вас, що це зовсім не так.

Насправді все набагато простіше. Для практичних цілей зовсім не потрібно завжди задавати питання: “А як же все це реалізується в природі?” (Богом, якщо хочете, або природною еволюцією). Можна запитати і так, як це зроблено в заголовку цієї статті: “А як має бути влаштований людський мозок, щоб його можна було б відтворити і реалізувати в машині?” Така постановка питання, безумовно, є еретичною. І я не збираюся вдавати, що я не розумію цього.

До того ж, і як вже було заявлено вище, я зовсім не збираюся вдавати, що я “Biologically inspired” (натхнений біологією), що я знаю або здогадуюся, як усе це влаштовано в природі. Зовсім навпаки, я стверджую, що я цього не знаю і знати не хочу. Я стверджую, що природа не інженер, а значить ніколи не придумує і не винаходить нічого нового. Вона пристосовує і підганяє те, що у неї вже є під рукою. Іноді з цього виходять чудові речі. Але завжди на це йде маса сил і часу, мільйони і мільярди років, яких у нашому розпорядженні немає.

При цьому, найголовніші речі, які визначили і визначають долю *Homo sapiens* на цій планеті, виникли не в результаті природної еволюції, а були саме винайдені, придумані на порожньому місці самою людиною, і природних аналогів у них немає. Хочете приклади? Будь ласка: перш за все, це цілеспрямована праця і знаряддя праці. Потім використання вогню для приготування їжі (і взагалі, використання вареної їжі). Потім — це наша розмовна мова, а потім вже і рахунок, і письмо, у свою чергу. А далі вже —

маса всяких інших винаходів, які тепер називаємо культурою, і без яких сьогодні сучасної людини вже неможливо собі уявити.

Проте, я хотів би повернутися і все викласти по порядку, не роблячи своїх висновків зі спостережуваних у природі фактів, а відшукуючи в масі відомих уже експериментальних даних підтвердження своїм міркуванням, які йдуть “згори-вниз”. При цьому я виходжу з наступного припущення: якщо справді висловлювані мною висновки не є якимись довільними вигадками, а мають загальний і фундаментальний характер, то ці явища повинні спостерігатися в природі і бути присутніми на всіх рівнях еволюційного розвитку живих істот від мікроорганізмів і бактерій до вищих приматів і людини.

Отже, **перше твердження: інформація є лінгвістичним описом.**

Сучасні уявлення про генетичний код дуже добре узгоджуються з цим твердженням. Що нейрони в мозку обмінюються між собою пачками електричних імпульсів, які є інформацією, здається, суперечать цьому. Але це тільки здається. Розповідь про астроцити і безімпульсний обмін інформацією між нейронами й астроцитами я вже наводив раніше, так само, як і міркування про відсутність зв'язку між нейронними імпульсами й інформацією. Повторюватися не буду. Біологам доведеться переглянути свої погляди з цього приводу.

Без розуміння інформації як лінгвістичного опису неможливо зрозуміти різницю між фізичною і семантичною інформацією. *Фізична інформація* — це описи стану зовнішнього світу, які потрапляють до нас із зовнішнього світу через наші органи чуття. *Семантична інформація* — це та інтерпретація, яку отримує у нас фізична інформація, це та розповідь з минулого життя (свого або чужого), в яку фізична інформація, що надійшла, прибудовується, знаходячи таким чином сенс і підставу для ухвалення подальших рішень.

Для того, щоб мати можливість аналізувати (прибудовувати) фізичну інформацію, що знову надходить, і відповідним чином реагувати на неї, система (і будь-який живий організм, зокрема) повинні володіти пам'яттю, де мають зберігатися всі ці потрібні розповіді з минулого життя. При цьому за аналогією з добре відомими нам комп'ютерами я смію стверджувати, що *будь-яка жива система має володіти двома видами пам'яті — жорсткою (або, як в комп'ютерах це називають, *hardwired*) і гнучкою (або, як в*

*комп'ютерах це називають, *softwired*) пам'яттю*. Тобто, наявність *жорсткої і гнучкої пам'яті* є обов'язковою умовою ефективної роботи будь-якої розумної системи. (У 6-му розділі я називав ці два види Генетичною й Оперативною пам'яттю. За змістом це одне і те саме, тому я і далі користуватимуся цими термінами почергово).

Сказавши це, потрібно негайно подивитися, а що ж ми маємо в цьому сенсі в біології. З людиною тут усе зрозуміло — наш генетичний код, що зберігається століттями, цілком підходить під визначення жорсткої пам'яті. Пам'ять же, яка є у нас в голові, і якою ми вдень і вночі користуємося, добре укладається в поняття гнучкої пам'яті. А що ж ми маємо в інших, нижчого рівня біологічних системах? У бактерій, наприклад?

Я був здивований, як багато вже зроблено в цьому сенсі (і досліджено, і опубліковано) в галузі обробки інформації у бактерій. Ізраїльтяни, виявляється, йдуть тут у перших рядах. Я наводжу в списку джерел декілька опублікованих зовсім недавно чудових робіт *Бен-Якова* [3-6] і *Яблонкі* [7]. Деякі з них ще в процесі друку, і можуть бути отримані з сайту *Ешеля Бен-Якова* (<http://star.tau.ac.il/~eshel/>). (Звичайно, і всі інші цитовані роботи можна знайти там само). Чудові ці статті тим, що попри традиційні підходи й уявлення (які неймовірно утрудняють правильне тлумачення спостережуваних явищ), їх автори зуміли, проте, правильно описати побачене.

Далі можна сторінками цитувати з *Бен-Якова*, але я відберу лише найістотніше. Вже у бактерій справді можна спостерігати роздільне існування жорсткої і гнучкої пам'яті. *Жорстка пам'ять* — у *хромосомах основного генома*, *гнучка пам'ять* — у *плазмідах*. За Вікіпедією, плазміді — це фрагменти ДНК, розташовані в клітинах поза хромосомами. Можуть вбудовуватися в основний геном, можуть і “вирізатися” з основного генома й існувати (співіснувати з ним), а також і поза ним. Бен-Яков вважає їх основним елементом, відповідальним за зв'язок між бактеріями (як усередині однієї колонії, так і між різними колоніями) і за здійснюване при цьому перенесення інформації (Ben-Jacob, 2008). За Бен-Яковом саме так відбувається *Горизонтальне Перенесення генів у бактерій*, тобто копіювання з плазміда однієї клітини в плазмід іншої, а потім вбудовування цього плазміда в основний геном, і вже потім передача новоутвореної інформації від батьків до дітей традиційним шляхом, так званім шляхом *Вертикального перенесення*

генетичної інформації. Саме так (за Бен-Яковом) і виробляється у бактерій резистивність до антибіотиків (Ben-Jacob et al., 2004).

Проте велика частина інформації, перенесеної горизонтально між бактеріями, не стає частиною генетичного коду, а залишається в Оперативній пам'яті кожної бактерії (у її плазмідах). Це веде до утворення так званої *Колективної пам'яті* в колонії бактерій і явних елементів соціальної поведінки у бактерій, що вимагає наявності загальної семантики і постійного спілкування між членами співтовариства (Ben-Jacob, 2008). Все це неспростовно свідчить про те, що 1) у бактерій існує Оперативна пам'ять, 2) зберігається в ній семантична інформація, без якої неможливі осмислені колективні дії, 3) ця семантична інформація є продуктом спілкування й обміну інформацією між клітинами, і 4) ця інформація записана визначеною, всім клітинам зрозумілою мовою. Бен-Яков відзначає навіть існування окремих діалектів цієї мови і описує, як за загальним договором бактерії можуть змінити наявний діалект, щоб захистити себе від бактерій-відступників, що паразитують на працях співтовариства (Ben-Jacob, Shapira & Tauber, 2006).

Зрозуміло, що існування оперативної пам'яті, відособленої і фізично відокремленої від генетичної пам'яті, безперечно, мало представляти еволюційну перевагу, оскільки забезпечувало велику гнучкість і пристосовність до умов швидкозмінного зовнішнього середовища. Генетична пам'ять повинна акумулювати більш довгостроковий досвід. Вона зобов'язана бути консервативнішою, щоб забезпечувати стійкість виду. Оперативна ж пам'ять потрібна для короткочасної адаптації, хоча з часом вона може переходити в стійку генетичну пам'ять (якщо умови життя, що змінилися, достатньо довго зберігалися в часі).

Зрозуміло, що в процесі еволюції всі елементи Оперативної пам'яті та її взаємодії з Генетичною пам'яттю постійно удосконалювалися й ускладнювалися, досягнувши своєї межі та досконалості в сьогоденній людській пам'яті. Вже було показано, що навіть на рівні бактерій семантична інформація надходить в оперативну пам'ять ззовні. А це, у свою чергу, припускає наявність форм взаємозв'язку і спілкування, а також обов'язково наявності мови такого спілкування. Хай дуже примітивної, хай навіть рудиментарної мови (описи і спілкування), але обов'язково мови. Тільки цією рудиментарною мовою окремі клітини і

цілі співтовариства клітин можуть обмінюватися між собою інформацією. Це має бути особливо важливим і вимагає постійного підкреслення, коли йдеться про таке спеціальне співтовариство клітин як мозок, яке унікальне в тому сенсі, що призначене тільки для обробки інформації, що надходить від інших клітинних співтовариств (від органів чуття, наприклад, живого організму). Природно, що обмін і накопичення інформації в оперативній пам'яті, обмежені рамками обміну всередині одного живого організму, завжди будуть дуже вбогими і мізерними. Винахід розмовної мови, а услід за ним писемності та рахунку, надзвичайно прискорили й інтенсифікували процес обміну і накопичення інформації в оперативній пам'яті *Homo sapiens*. Але його базові, фундаментальні принципи, безперечно, залишалися незмінними. У цьому сенсі цікаво було б уявити собі, як реалізуються в нас у мозку ці процеси запису і читання до/з оперативної пам'яті. Оскільки біологи про це нічого не знають, ми можемо дозволити собі знову поміркувати логічно. Блок-схема і передбачуваний алгоритм процесу обробки семантичної інформації, які ми описали раніше, нічого не говорять про те, як і звідки вноситься до цього блоку обробки ієрархічне представлення семантичної інформації. Цим-то ми зараз і займемося.

Оскільки оперативна інформація повинна представлятися текстовими стрінгами, розташованими поза основним геномом, природно припустити, що ці стрінги зберігаються в так званих "дендритних шипиках" на дендритних гілках нейрона і в синаптичних капсулах на його аксонних відростках. Інших можливих місць зберігання цієї інформації я не знаю. (Нещодавно, 17 грудня 2009 р., у журналі "Nature" була опублікована стаття (Yang et al., 2009), в якій описується розміщення пам'ятних записів у дендритних шипиках, що спостерігалися *in vivo* на підслідних мишах. Яке чудове і своєчасне підтвердження моїх спекуляцій!).

Оскільки на рівні бактерій інформація, що міститься в оперативній пам'яті (у плазміді), і генетична інформація в хромосомах основного генома легко рекомбінують, природно припустити, що вони записані однією й тією самою мовою. Так само природно припустити, що в процесі еволюції ця мова видозмінювалася. А тому в системі завжди існували й існують не одна, а безліч мов опису, так само, як і правила перекладу (переходу) з однієї мови на іншу. Підтверд-

женням цьому можуть служити добре відомі процеси перекристалізації записів короткострокової пам'яті в довгострокові записи, так звана *Довготривала Потенціалізація (Long-Term Potentiation)*. Незмінними при цьому мали би зберігатися лише форма і структура запису, тобто для кожного запису в оперативну пам'ять мала би бути синтезована довга послідовність молекулярних кодованих елементів у вигляді подвійної спіралі ДНК. Мені не дуже зрозуміло поки, як саме синтезується цей первинний запис, але мені здається, що я добре розумію, як повинне відбуватися зчитування інформації з операційної пам'яті. Я стверджую, що це мусить відбуватися так само, як це відбувається при реплікації генів у генетичній пам'яті. Тобто подвійна спіраль розплітається, одна частина відновлюється до свого первинного вигляду і зберігається на місці (для подальшого використання), а друга частина (також відновлена) "набирає ходу", тобто переноситься в ту ділянку мозку, де фактично здійснюється процес пошуку оптимального поєднання вхідної інформації і відповідної їй описаної розповіді, що зберігається в операційній пам'яті.

Природно припустити, що при відновленні копійованої половинки запису, ще до того, як вона "набирає ходу", відбувається попутне "редагування" цього запису (точно так, як це відбувається при реплікації генетичної інформації), тобто за якимись невідомими для нас правилами, відбувається заміна і перестановка окремих букв, слів, цілих фраз і навіть окремих шматків тексту в первинному записі (Уотсон, 2008). Колись це називали мутаціями. Сьогодні це називають редакціями.

Таким чином те, що колись було ілюзорним процесом мислення, сьогодні матеріалізується в простий і цілком зрозумілий алгоритм обробки інформаційних текстів.

8. Епілог

Пора закінчувати. Я, як умів, виклав і подав тут з усіма подробицями розповідь про те, куди мене завело Колмогоровське розуміння інформації як лінгвістичного опису (або, коротше кажучи, текстових стрінгів). Залишилася одна маленька деталь, один філософський нюанс всієї цієї історії, який мені хотілося б викласти хоч під завесу.

Оперативна інформація (ієрархія семантичної та фізичної інформації, записана у формі розповіді) займає сьогодні таке важливе місце в розвитку виду *Homo sapiens*, що не гріх

було б замислитися: А чи не закінчилася для нас дарвінівська еволюція, поступившись своїм місцем Культурній еволюції, в якій ми тепер живемо і яку навколо себе спостерігаємо?

Обмежені можливості накопичення інформації в Оперативній пам'яті еволюційно нижчих живих істот змінилися вибуховим зростанням можливостей людського розуму, коли з винаходом розмовної мови, писемності і рахунку стало звичайним і доступним поповнення оперативної пам'яті окремого індивідуума з резервуару зовнішньої колективної пам'яті (спочатку найближчого роду і племені, територіально відособленого народу, а потім, у міру винаходу нових засобів зберігання і перенесення інформації, розширення потенційного резервуару зовнішньої колективної пам'яті до розмірів усього людства).

Доступність інформації із зовнішньої колективної пам'яті і можливість переважання її в оперативній пам'яті окремого індивіда, а також еволюційні переваги такого доступу сьогодні настільки очевидні, що немає сенсу ще докладніше зупинятися на цьому. Потрібно тільки відзначити, що, у свою чергу, все це веде до вибухового розширення обсягу накопиченої колективної пам'яті людства, до стрімкої інтенсифікації можливостей колективного людського розуму.

Одним із аспектів цієї інтенсифікації можливостей є те, що ми (людство) дістали можливість втручатися в самі результати природної еволюції — втручатися в генетичну пам'ять, свою і всіх інших живих істот.

Вчора оголосили про присудження Нобелівської премії дослідникам механізму діяльності телломерів. Генетично модифіковані продукти (Європа, щоправда, ще відмовляється їх їсти, але в Індії і Китаї це питання вже давно не на порядку денному), дослідження і лікування різних генетично зумовлених захворювань, численні приклади генної інженерії, — все це свідочить про те, як далеко ми вже пішли від часу дарвінівської еволюції, і як впливає тепер на наше життя культурна еволюція, якій ми зобов'язані сьогодні буквально всім.

І ще про одну річ я хотів би згадати: *мем — еквівалент гена, відповідальний за передачу культурної інформації*, придуманий *Докінзом* (Докінз, 1976). Коли *Чарльз Дарвін* 1859 р. завершував писати своє "Походження видів", про механізми спадковості ще й не згадувалося. Перші закони спадковості були відкриті лише 1865 р. Тут же вони були намірно забуті, щоб бути знов перевинайденими через тридцять п'ять років, 1900 р. Сам термін "ген" був введений в обіг тільки 1909 р.

У 1953 році *Уотсон* і *Крік* відкрили двоспіральну структуру молекули ДНК — фізико-хімічну реалізацію живого гена. Таким чином, майже сто років теорія еволюції існувала і розвивалася незалежно від нових ідей і понять, пов'язаних з дослідженням її основних матеріальних носіїв. Тому, коли в середині 1970-х років *Річард Докінз* зробив спробу доповнити *Дарвіна* новітніми відкриттями з галузі генетики, його книгу чекав приголомшливий успіх. Вона одразу стала бестселером,

витримала декілька видань, найостанніше з яких з'явилося 2006 р. Проте, попри всю оригінальність і новизну підходу, Докінз не міг вирішити одну з основних проблем еволюційного розвитку — *як зв'язані між собою генотип і фенотип, як потрапляє до нас із нашого оточення культурна інформація?* Не заглиблюючись дуже далеко в подробиці, скажу лише, що для вирішення цієї проблеми Докінз винайшов і ввів в обіг поняття "мема" — еквівалента і подібності гена, відповідального за передачу фенотипічних та інших культурних особливостей в процесі еволюції конкретного організму. Ідея була вмить підхоплена, і ціла нова галузь знань, що отримала назву "Меметика", почала стрімко розвиватися (Heylighen & Chielens, 2008).

Ні Докінз, ні його численні послідовники не хочуть помічати, як далеко завело (і відвело убік) їх геноцентричне розуміння еволюції (тобто, що за все те, що відбувається з нами несуть відповідальність тільки наші гени). Адже ген — це всього лише довгострокова (жорстка) пам'ять із записаною в ній історично незмінною інформацією. А те, що Докінз називає мемом, — це всього лиш оперативна (гнучка) пам'ять, в яку переважання колективної пам'яті людства те знання, яке певному конкретному індивіду в певний конкретний момент забезпечує найоптимальніші умови існування і виживання.

І все це ми вже обговорювали вище. І немає сенсу повертатися до цього знову і знову.

Переклад Олени Засць

Література

(Щоб не перевантажувати занадто довгим списком джерел, я наводжу нижче посилання на ті статті, які вперше трапилися лише в цьому тексті або становлять особливий інтерес. Всі інші посилання можна знайти в моїх раніше опублікованих статтях, котрі доступні на сайті <http://www.vidia-mant.info>).

1. **Torralba A.** (2009). How many pixels make an image? // Visual Neuroscience. — Vol. 26, Issue 1. — p. 123-131. Available: <http://web.mit.edu/torralba/www/>.
2. **Treisman A. & Gelade, G.** (1980). A feature-integration theory of attention // Cognitive Psychology. — Vol. 12, pp. 97-136.
3. **Ben-Jacob E.** (2008). Social behavior of bacteria: from physics to complex organization // The European Physical Journal B. — vol. 65, pp.315-322.
4. **Ben-Jacob E., Shapira Y. & Tauber A.** (2006). Seeking the foundations of cognition in bacteria: From Schrodinger's negative entropy to latent information // Physica A. — vol. 359. — pp. 495-524.
5. **Ben-Jacob E. & Levine H.** (2006). Self-engineering capabilities of bacteria // Journal of The Royal Society Interface. — vol. 3, № 6. — pp. 197-214.
6. **Ben-Jacob E., Becker I., Shapira Y. & Levine H.** (2004). Bacterial linguistic communication and social intelligence // Trends in Microbiology. — vol. 12, № 8. — pp. 366-372.
7. **Jablunka E. & Ginsburg S.** (2008). Epigenetic learning in non-neural organisms // Journal of Bioscience. — vol. 33, № 4. — pp. XX-YY.
8. **Heylighen F. & Chielens K.** (2008). Cultural Evolution and Memetics. In: B. Mayers (Ed.), Encyclopedia of Complexity and System Science, Springer-Verlag, Berlin Heidelberg, 2008. Available (Доступно на сайті): <http://pespmc1.vub.ac.be/papers/PapersFH2.html>.

9. **Diamant E.** (2005). Searching for image information content, its discovery, extraction, and representation // Journal of Electronic Imaging. — vol. 14, № 1.

10. **Diamant E.** (2005a). Does a plane imitate a bird? Does computer vision have to follow biological paradigms?, In: De Gregorio, M., et al, (Eds.), Brain, Vision, and Artificial Intelligence, First International Symposium Proceedings. LNCS, vol. 3704, Springer-Verlag, pp. 108-115, 2005. Available: <http://www.vidia-mant.info>.

11. **Diamant E.** (2008). Unveiling the mystery of visual information processing in human brain // Brain Research. — vol. 1225, pp. 171-178.

12. **Ванюшин Б.Ф.** (2004). Матеріалізація епігенетики, или небольшие изменения с большими последствиями. Доступно: <http://wsyachina.narod.ru/chemistry/apygenetics.html>.

13. **Ренун В.С.** (2000). "Ковчег Жизни" на стапелях эволюции // Новый Мир. — № 12.

14. **Уотсон Дж.** (2008). ДНК и мозг: в поисках генов психических заболеваний, Публичные лекции фонда "Династия", Публичная лекция фонда "Династия", 3 июля 2008 года, Москва, Дом ученых, Доступно: <http://elementy.ru/lib/lectons>.

15. **Гельфанд И.М., Розенфельд Б.И., Шифрин М.А.** (1989). Очерки о совместной работе математиков и врачей. Издательство АН СССР Наука, Москва, 1989.

16. **Докінз Р.** (1976). Эгоистичный ген. Доступно: http://fictionbook.ru/author/dokiniz_richard/.

17. **Yang G., Pan F., and Gan W-B.** (2009). Stably maintained dendritic spines are associated with lifelong memories // Nature. — vol. 462. — p.920-924.