



Визитка

ВЛАДИМИР БРЮКОВ, НЕЗАВИСИМЫЙ АНАЛИТИК

Лингва франка сети 2036 года

Как изменятся языковые предпочтения пользователей

Колыбель Всемирной сети — США, потому доминирование англоязычных сайтов характерно для Интернета с момента его рождения. И сегодня здесь язык великого Шекспира продолжает сохранять свое неоспоримое лидерство

Лингвистические диспропорции современного Интернета

Согласно данным Википедии 51,3% интернет-трафика (просмотра страниц этой мультиязычной энциклопедии) за период с ноября 2009-го по октябрь 2010 года приходится на англоязычный раздел Википедии (en.wikipedia.org). В то же время общее число владеющих английским в качестве родного или второго языка составляет 1500 млн человек или 22,2% от населения мира. Следовательно, индекс обеспеченности Интернетом для владеющих английским языком составляет 231,1% (доля в интернет-трафике просмотра страниц на английском языке/доля англоязычного населения

в мире *100%), т. е. в 2,31 раза выше, чем для среднестатистического жителя планеты.

В качестве репрезентативной выборки для анализа языковых предпочтений интернет-пользователей мы взяли данные Википедии по доле интернет-трафика на различных языках за период с ноября 2009-го по октябрь 2010 года. Этот выбор сделан не только потому, что ежемесячно в мультиязычной энциклопедии просматривается около 10,7 млрд страниц, но и в силу того, что в этом проекте широко представлены интернет-пользователи разных стран и языковых сообществ (см. данные о посещении пользователями Википедии на сайте <http://stats.wikimedia.org>).

По данным Википедии, еще более высокий индекс обеспеченности наблюдается в японоязычном (399,3%) и немецкоязычном Интернете (284,9%), на которые приходится по 7,8% всего трафика, в то время как доля владеющих японским и немецким языком составляет лишь 2,7 и 2,0% от населения планеты. Вместе с тем индекс обеспеченности Интернетом для китайязычного (7,3%), арабоязычного (3,8%) и хиндиязычного (0,1%) населения и для владеющих рядом других распространенных восточных языков сегодня в десятки — в тысячу раз ниже по сравнению с обеспеченностью тех, кто владеет, например, английским, немецким или японским языками (см. таблицу 1). Так, на долю тех, кто владеет китайским (упрощенное письмо) и арабским языком, а также языком хинди приходится 19,2; 7,8 и 8,1% населения мира, но доля трафика на этих языках составляет лишь 1,4; 0,3 и 0,01% (см. таблицу 1).

Очевидно, что более низкая доля ряда мировых языков в интернет-трафике объясняется объективными факторами. Во-первых, более низким уровнем компьютеризации того или иного языкового сообщества; во-вторых, низкой долей числа интернет-пользователей; в-третьих, низкой скоростью работы Интернета либо нехваткой средств для его оплаты у населения развивающихся стран; в-четвертых, дефицитом в Интернете информации на данном языке либо наличием цензуры (как, например, в КНР).

Вместе с тем бурные темпы подключения к сети (наблюдаются как на развитых, так и особенно на развивающихся

Таблица 1. Мировые языки и обеспеченность интернет-трафиком владеющих этими языками

Язык	Количество владеющих языком (как родным или вторым), млн чел.	Доля, % (от населения мира)	Доля интернет-трафика, %	Индекс обеспеченности интернет-трафиком, %
Английский	1500	22,2	51,3	231,1
Китайский	1300	19,2	1,4	7,3
Хинди	550	8,1	0,01	0,1
Арабский	530	7,8	0,3	3,8
Испанский	500	7,4	7,2	97,3
Малайский	300	4,4	0,04	0,9
Португальский	290	4,3	2,3	53,6
Русский	278	4,1	3,3	80,2
Индонезийский	250	3,7	0,30	8,1
Бенгальский	230	3,4	0,003	0,1
Французский	200	3,0	4,2	141,9
Немецкий	185	2,7	7,8	284,9
Японский	132	2,0	7,8	399,3

Источник: расчеты автора по данным Википедии



Если прогноз сбудется,
почти **60% всех**
интернет-сайтов в 2036 году
станут англоязычными

рынках) делают перспективу всеобщей интернетизации вполне достижимой в ближайшие два-три десятка лет. Так, по данным сайта <http://www.internetworldstats.com>, с 2000-го по 2010 год количество арабоязычных, русскоязычных, китайоязычных и португалоязычных интернет-пользователей увеличилось соответственно на 2501,2; 1825,8; 1277,4 и 989,6%, в то время как количество англоязычных, немецкоязычных и японоязычных интернет-пользователей увеличилось лишь на 281,2; 173,1 и 110,6%.

Когда закончится всеобщая интернетизация?

Вспользуемся данными сайта www.internetworldstats.com по росту доли интернет-пользователей к населению мира за период с декабря 1995-го по сентябрь 2010 года для того, чтобы определить примерную дату завершения всеобщей интернетизации. При этом мы использовали для аппроксимации тренда три вида статистических моделей: линейный тренд, тренд в виде полинома второй степени и тренд в виде логарифмической функции. В результате мы получили три уравнения регрессии, все коэффициенты которых оказались статистически значимыми с 95% уровнем надежности (см. таблицу 2).

В статистической литературе рекомендуется делать прогнозы по моделям, у которых относительная ошибка аппроксимации не превышает 7%. Поскольку относительная ошибка аппроксимации у линейного тренда оказалась существенно

выше этой цифры (см. таблицу 2), мы решили отказаться от использования этой модели для прогнозирования.

Вместе с тем, судя по коэффициенту детерминации и относительной ошибке аппроксимации, прогностические возможности функции в виде полинома второй степени оказались вполне сопоставимы с аналогичными возможностями логарифмической функции. Более того, если бы нашей целью являлось составление прогноза на ближайшие один-два года, то вполне можно было бы воспользоваться полиномом второй степени. Однако в данном случае перед нами стоит совершенно иная задача: рассчитать дату завершения всеобщей интернетизации. В этом случае для прогноза лучше подходит логарифмическая модель, т.к. в силу ее специфики эта модель на первом этапе показывает стремительный рост доли интернет-пользователей, который затем будет постепенно замедляться по мере насыщения рынка интернет-услугами. Следует заметить, что такая динамика характерна для продвижения любого товара на рынке, поэтому прогноз по логарифмической модели станет у нас основным. Результаты прогнозирования по логарифмической модели (выделены жирным шрифтом) занесены в таблицу 3, рядом с которыми для сравнения помещены прогнозы, сделанные по двум другим моделям.

Судя по таблице 3, 100% уровень интернетизации планеты по линейному тренду может быть достигнут к середине 2048 года, по полиному второй степени – к середине 2027-го,

Таблица 2. Качество статистических моделей по аппроксимации роста доли интернет-пользователей

Параметры	Линейная функция	Функция в виде полинома второй степени	Логарифмическая функция
Формула	$Y = -0,0360 + 0,00165t$	$Y = 0,000635t + 0,0000054t^2$	$Y = -8,2537 + 1,336\ln(t)$
Коэффициент детерминации R^2	0,960	0,9912	0,9905
Относительная ошибка аппроксимации, %	15,3	6,4	5,7

Зависимая переменная Y – доля интернет-пользователей (к населению мира); независимая переменная t – время (порядковое значение декабря 1995 года $t=1$).
Источник: расчеты автора по данным сайта <http://www.internetworldstats.com>

по логарифмической функции – к середине 2036 го. С учетом специфики используемых моделей возьмем в качестве основной последнюю дату. При этом мы исходим из предположения, что рост интернетизации будет происходить быстрее, чем рост населения Земли, как это имело место в период с 1995-го по 2010 год.

Прогноз сроком на 25 лет невозможен без определенных сценарных допущений. Во-первых, предположим, что доля языковых сообществ в мировом населении в этот период не претерпит существенных изменений. Во-вторых, будем исходить из того, что разница в среднем интернет-трафике пользователей, работающих в развитых и развивающихся странах, в течение ближайших 25 лет должна серьезно сократиться либо вообще исчезнуть. Тренд на выравнивание среднего трафика пусть будет обусловлен удешевлением интернет-услуг, повышением доступности необходимой

для этих целей компьютерной техники, а также более быстрым ростом благосостояния в развивающихся странах.

Согласно данным Википедии за период с ноября 2009-го по октябрь 2010 года на каждого интернет-пользователя из развитых стран Севера в среднем приходилось по 11 просмотренных страниц в месяц. В то время как пользователь из развивающихся стран Юга в среднем просматривал по две страницы, т.е. интернет-трафик последнего составлял лишь 18,2% от среднего трафика первого.

Если предположить, что к 2036 году разница в среднем трафике между пользователями из развитых и развивающихся стран полностью исчезнет, а уровень интернетизации к тому времени достигнет 100%, то доля национального интернет-трафика асимптотически будет стремиться к доле населения данной страны в населении мира.

При долгосрочном прогнозе приходится исходить из нескольких сценариев развития. Поэтому сценарий, предполагающий 100% устранение различий между странами Севера и Юга в величине среднего интернет-трафика, приходящегося на пользователя, назовем оптимистическим. В свою очередь, будем иметь в виду также консервативный и оптимальный сценарии развития, которые предполагают, что к 2036 году эта разница сократится либо на 50, либо на 75%.

Прогнозы по изменению интернет-трафика для языковых сообществ на 2036 год были составлены следующим образом. Во-первых, мы нашли индексы роста трафика для каждой страны согласно оптимистическому сценарию развития. При этом индексы вычислялись путем деления доли этих стран в населении мира на их долю в интернет-трафике в 2010-м. Во-вторых, для того чтобы рассчитать индексы роста трафика для каждой страны в случае реализации консервативного либо оптимального сценария, национальные индексы роста трафика, рассчитанные по оптимистическому сценарию, корректировались с учетом 50 или 75% сокращения разницы в среднем трафике. В-третьих, данные Википедии дают не только долю каждого языка в мировом интернет-трафике, но и долю пользователей каждой страны в использовании данного языка (см. графу 2010 год в таблицах 5-11). Следовательно, умножив эти доли на индексы роста интернет-трафика в данной стране по различным сценариям, а затем суммировав полученные результаты по каждому языку, мы спрогнозировали изменение языковых предпочтений пользователей сети к 2036 году.

Языковые предпочтения сети через четверть века

В результате у нас получилась таблица 4, согласно которой к 2036 году в языковых предпочтениях пользователей интернета произойдут серьезные изменения. Во-первых, доля японоязычного и немецкоязычного Интернета должна резко сократиться, что обусловлено распространением этих языков в развитых странах мира, где темпы интернетизации сильно замедляются. В то время как второе и третье место по доле в интернет-трафике займут китайский (рост по оптимистическому сценарию на 8,7 процентных пункта) и испанский языки (снижение на 2,4 процентных пункта). Однако лидерство английского языка станет еще более прочным: согласно оптимистическому сценарию его доля в трафике увеличится с 51,3 до 59,0%. В целом же доля трех самых крупных языков Интернета в течение 25 лет должна вырасти с 66,9 до 73,9%.

Таблица 3. Завершение всеобщей интернетизации согласно прогнозу по трем статистическим моделям

Месяц	Доля интернетизации (от населения мира), %	Аппроксимация по линейной функции	Аппроксимация по полиному второй степени	Аппроксимация по логарифмической функции
Дек. 96	0,9	-1,5	0,9	0,8
Дек. 97	1,7	0,5	1,9	1,9
Дек. 98	3,6	2,5	3,1	3,2
Дек. 99	4,1	4,5	4,4	4,7
Дек. 00	5,8	6,5	5,9	6,3
Авг. 01	8,6	7,8	6,9	7,4
Сент. 02	9,4	9,9	8,8	9,4
Дек. 03	11,1	12,4	11,2	11,7
Дек. 04	12,7	14,4	13,3	13,7
Дек. 05	15,7	16,4	15,5	15,7
Дек. 06	16,7	18,3	17,9	17,9
Дек. 07	20,0	20,3	20,5	20,0
Дек. 08	23,5	22,3	23,2	22,3
Дек. 09	26,6	24,3	26,1	24,6
Июнь 10	28,7	24,8	26,8	25,2
Сент. 10	28,8	25,3	27,6	25,8
Дек. 11	-	27,7	31,5	28,8
Дек. 12	-	29,7	34,7	31,2
Дек. 13	-	31,7	38,2	33,7
Дек. 14	-	33,7	41,8	36,3
Дек. 15	-	35,6	45,5	38,9
Дек. 20	-	45,5	66,6	52,5
Янв. 25	-	53,6	86,7	64,3
Июль 27	-	58,6	100,3	71,8
Янв. 30	-	63,5	114,8	79,6
Май 36	-	76,0	155,9	100,0
Янв. 40	-	83,3	182,6	112,4
Июнь 48	-	100,0	251,7	142,0

Источник: расчеты автора по данным сайта <http://www.internetworldstats.com>

В таблице 5 (смотрите на сайте журнала) можно увидеть, за счет каких стран должен к 2036 году вырасти интернет-трафик английского языка. Согласно оптимистичному сценарию доля Индии в англоязычном трафике вырастет с 3,5 до 27,7 %, Китая – с 0,5 до 13,3%, других стран (по большей части бывших британских колоний) – с 10,7 до 37,7%. Очевидно, что рост англоязычного Интернета будет обусловлен высокими темпами интернетизации развивающихся стран, в то время как доля США, Великобритании, Канады, Австралии и других развитых стран в англоязычном трафике существенно снизится.

Судя по таблице 6 (смотрите на samag.ru), доля Испании в испаноязычном интернет-трафике согласно оптимистичному сценарию сократится более чем на 10 процентных пунктов. Существенно уменьшится доля в этом трафике у Чили и у испаноязычных жителей США. Немного подрастет доля современного лидера – Мексики, на которую будет приходиться 28% всего испаноязычного трафика, но еще более серьезно нарастят долю такие латиноамериканские страны, как Венесуэла, Перу, Гватемала, Боливия.

Как следует из таблицы 7 (смотрите на сайте журнала), к 2036 году резко упадет доля Франции во франкоязычном интернет-трафике – почти на 40 процентных пунктов. Одновременно с этим значительно сократится доля Канады, Бельгии и Швейцарии. Зато существенно вырастет доля в этом трафике Алжира, Марокко, Туниса и других стран – бывших африканских колоний Франции.

Судя по таблице 8 (смотрите на samag.ru), доля России в русскоязычном трафике к 2036 году немного снизится. Сократится в этом трафике также доля русскоязычных пользователей из Германии, США, Израиля и Латвии, зато существенно вырастет доля Казахстана, Беларуси и Украины. Несмотря на попытки ряда национальных государств силовыми методами сократить сферу применения русского языка, его роль в постсоветском секторе Интернета в ближайшие 25 лет вряд ли уменьшится. В пользу этого говорит на порядок большая востребованность русского языка по сравнению с другими языками бывших республик СССР. Так, любая статья на русском языке в Википедии в декабре 2010 года в среднем имела 1,2 просмотра в час, в то время как для статьи на украинском языке эта цифра равнялась 0,15, на латвийском – 0,19 и на эстонском языке – 0,17.

Из таблицы 9 (смотрите на samag.ru) можно сделать вывод, что львиная доля португалоязычного интернет-трафика и в 2036 году будет приходиться на Бразилию и практически не изменится. Доля Португалии в этом трафике серьезно сократится – с 10,7 до 3,4%, однако это сокращение будет компенсировано за счет прироста доли других стран – в основном бывших африканских колоний Португалии.

Судя по таблице 10 (смотрите на samag.ru), доля континентального Китая в китайоязычном интернет-трафике к 2036 году должна вырасти колоссально – с 22,6 до 95,3%, т.е. на 72,7 процентного пункта. В то время как соответствующие доли Тайваня, Гонконга и китайской диаспоры в США должны резко сократиться. Например, доля Тайваня в ближайшие 25 лет уменьшится с 36,9 до 2,4%.

Из таблицы 11 (смотрите на samag.ru) нетрудно сделать вывод, что наибольшая доля в арабоязычном интернет-трафике в 2036 году будет приходиться на Египет, который займет по этому показателю первое место. Кроме того,

значительно вырастет доля в этом трафике Судана, Ирака и Йемена. Зато доля таких стран, как Саудовская Аравия (в 2010 году была на первом месте), Кувейт, а также доля арабов Израиля и арабской диаспоры в США и Германии значительно уменьшится.

Что в итоге?

В результате реализации прогноза почти 60% всех интернет-сайтов в 2036 году станут англоязычными, что окончательно закрепит за английским статус языка мирового интернет-общения. При этом сформируются три модели лингвистического поведения пользователей.

Модель первого типа будет характерна для тех пользователей, родным языком которых является английский, поиск в Интернете они будут вести преимущественно на своем родном языке, т.е. останутся монолингвами.

На модель второго типа перейдут большинство пользователей, родным языком которых является неанглийский. Для того чтобы максимизировать объем получаемой информации, они будут вынуждены стать билингвами, т.е. будут вести поиск не только на родном языке, но и на английском. Сейчас доля таких пользователей в неанглоязычных странах, как правило, не превышает 5-20%.

В некоторых странах, например в Алжире или Тунисе, доминирующее положение займет трилингвистическая языковая модель, т.е. наряду с арабским и французским языками большинство пользователей там перейдут еще и на английский. Аналогичная ситуация возникнет и в украинском Интернете, где наряду с украинским и русским большинство пользователей будут применять еще и английский язык. **БОЕ**

Таблица 4. Доля языков в мировом интернет-трафике

Язык	2010 год	2036 год			Рост, процентные пункты
		Консервативный сценарий	Оптимальный сценарий	Оптимистический сценарий	
Английский	51,3	55,1	57,1	59,0	7,7
Японский	7,8	4,8	3,4	1,9	-5,9
Немецкий	7,8	4,5	2,9	1,3	-6,5
Испанский	7,2	6,0	5,4	4,8	-2,4
Французский	4,2	3,3	2,9	2,4	-1,8
Русский	3,3	3,0	2,9	2,8	-0,5
Итальянский	2,8	1,8	1,3	0,8	-2,0
Португальский	2,3	2,4	2,4	2,4	0,1
Польский	2,2	1,3	0,9	0,5	-1,7
Китайский	1,4	5,8	7,9	10,1	8,7
Нидерландский	1,2	0,7	0,5	0,2	-1,0
Индонезийский	0,3	0,9	1,2	1,5	1,2
Арабский	0,3	0,7	0,9	1,1	0,8
Доля трех топ-языков	66,9	66,9	70,4	73,9	7,0

Источник: здесь и ниже расчеты автора по данным Википедии