

Волчек А.А., Парфомук С.И.

СРАВНИТЕЛЬНАЯ ОЦЕНКА МАРКОВСКИХ И НЕЛИНЕЙНЫХ МОДЕЛЕЙ ГОДОВОГО СТОКА РЕК БЕЛАРУСИ

Введение

В последнее время широко исследуется проблема многолетних колебаний речного стока. Исследования по созданию детерминистических моделей процессов формирования речного стока с помощью методов математической физики и их использованию в гидрологических прогнозах и расчетах, предложенные во второй половине прошлого столетия, не способны объяснить ни один эффект, характерный для многолетних колебаний годового стока [1]. Поэтому сегодня исследуются новые методы гидрологических прогнозов с применением мало-параметрических нелинейных динамических моделей речного стока, позволяющие описать физические механизмы цикличности многолетних колебаний речного стока.

Методика исследования

Пусть \bar{Q} – средний многолетний расход воды, а Q_t – расход воды в t -й момент времени. Тогда, приняв $X_t = (Q_t - \bar{Q})/\bar{Q}$, процесс многолетних колебаний стока можно описать с помощью следующего стохастического дифференциального уравнения Орштейна-Уленбека с непрерывным временем t [1]:

$$dX_t = -kX_t dt + \sigma dW_t, \quad (1)$$

где k^{-1} – время релаксации речного стока; σ – интенсивность «белого шума»; W_t – стандартный винеровский процесс.

Интенсивность «белого шума» определяется как

$$\sigma = C_v \sqrt{2k},$$

где C_v – коэффициент вариации речного стока, а коэффициент k из следующего соотношения:

$$k = -\ln r,$$

где r – автокорреляционная функция колебаний речного стока.

Уравнению (1) соответствует уравнение Фоккера-Планка-Колмогорова

$$\frac{\partial p}{\partial t} = \frac{\partial}{\partial x}(kxp) + \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2}, \quad -\infty < x < \infty, \quad (2)$$

а обратное уравнение имеет вид

$$\frac{\partial}{\partial t} p(x, t/y, 0) = -ky \frac{\partial}{\partial y} p(x, t/y, 0) + \frac{1}{2} \sigma^2 \frac{\partial^2 p(x, t/y, 0)}{\partial y^2}, \quad -\infty < x < \infty, \quad (3)$$

т. е. случайные колебания стока однородны во времени, а, значит, имеет место соотношение

$$p(x, t/y, 0) = p(x, 0/y, t).$$

Рассмотрим следующую задачу стохастической гидрологии. Пусть в начальный момент времени $t=0$ сток равен Q , а Q_* – некоторое фиксированное значение стока. Требу-

ется определить период времени, в течение которого значение стока будет находиться в пределах $[Q_*, \infty)$. Пусть T – момент времени, когда значение стока покинет полуинтервал $[Q_*, \infty)$. Тогда

$$\text{prob}(T \geq t) = G(Q, t) \quad G(Q, t) = \int_Q^\infty p(x, t/y, 0) dx.$$

Интегрируя (3) от Q_* до ∞ по x , получим

$$\frac{\partial G(Q, t)}{\partial t} = -kQ \frac{\partial G(Q, t)}{\partial Q} + \frac{\sigma^2}{2} \frac{\partial^2 G(Q, t)}{\partial Q^2}.$$

Граничные условия определяются исходя из поглощения значения функции при $Q=Q_*$, а также из отражения на бесконечности, т. е.

$$G(Q, t)|_{Q=Q_*} = 0, \quad \left. \frac{\partial G(Q, t)}{\partial Q} \right|_{Q=\infty} = 0.$$

Среднее время достижения границы Q_* определяется следующим соотношением:

$$T_1 = - \int_0^\infty t \frac{\partial G(Q, t)}{\partial t} dt = \int_0^\infty G(Q, t) dt.$$

Интегрируя (3) по t от 0 до ∞ и учитывая, что

$$\int_0^\infty \frac{\partial G}{\partial t} dt = G(x, \infty) - G(x, 0) = -1,$$

получим следующее уравнение для T_1 :

$$\frac{1}{2} \sigma^2 \frac{d^2 T_1}{dQ^2} - kQ \frac{dT_1}{dQ} = -1$$

при $\left. \frac{dT_1}{dQ} \right|_{Q=\infty} = 0, \quad T_1(Q)|_{Q=Q_*} = 0.$

Введя безразмерные величины

$$\theta_1 = kT_1, \quad \xi = Q \sqrt{\frac{2k}{\sigma^2}} = \frac{Q}{C_v},$$

$$\xi_* = Q_* \sqrt{\frac{2k}{\sigma^2}} = \frac{Q_*}{C_v}$$

получим

$$\frac{d^2 \theta_1}{d\xi^2} - \xi \frac{d\theta_1}{d\xi} = -1, \quad \left. \frac{d\theta_1}{d\xi} \right|_{\xi=\infty} = 0,$$

$$\theta_1(\xi)|_{\xi=\xi_*} = 0. \quad (4)$$

Таблица 1. Решения уравнения (4)

ξ^*	ξ											
	-2,5	-2	-1,5	-1	-0,5	0	0,5	1	1,5	2	2,5	3
-3	59,9	76,5	82,3	84,8	86,1	86,9	87,5	87,8	88,1	88,4	88,6	88,7
-2,5		16,6	22,4	24,9	26,2	27,0	27,5	27,9	28,2	28,4	28,6	28,8
-2			5,8	8,3	9,6	10,4	10,9	11,3	11,6	11,8	12,0	12,2
-1,5				2,5	3,8	4,6	5,1	5,5	5,8	6,0	6,2	6,4
-1					1,3	2,1	2,6	3,0	3,3	3,5	3,7	3,9
-0,5						0,7	1,3	1,7	2,0	2,2	2,4	2,6
0							0,5	0,9	1,2	1,4	1,6	1,8

Таблица 2. Решения уравнения (5)

ξ^*	ξ											
	2,5	2	1,5	1	0,5	0	-0,5	-1	-1,5	-2	-2,5	-3
3	59,9	76,5	82,3	84,8	86,1	86,9	87,5	87,8	88,1	88,4	88,6	88,7
2,5		16,6	22,4	24,9	26,2	27,0	27,5	27,9	28,2	28,4	28,6	28,8
2			5,8	8,3	9,6	10,4	10,9	11,3	11,6	11,8	12,0	12,2
1,5				2,5	3,8	4,6	5,1	5,5	5,8	6,0	6,2	6,4
1					1,3	2,1	2,6	3,0	3,3	3,5	3,7	3,9
0,5						0,7	1,3	1,7	2,0	2,2	2,4	2,6
0							0,5	0,9	1,2	1,4	1,6	1,8

Интегрируя систему (4) численным методом [2, 3], получили результаты, приведенные в табл. 1.

Пусть теперь в начальный момент времени $t = 0$ сток равен Q , а Q_* – также некоторое фиксированное значение стока, но уже большее исходного. Для определения периода времени, в течение которого значение стока будет находиться в пределах $(\infty, Q_*]$, используется система (4) с измененными начальными условиями, т. е.

$$G(Q, t)|_{Q=Q_*} = 0, \quad \left. \frac{\partial G(Q, t)}{\partial Q} \right|_{Q=-\infty} = 0.$$

Тогда будем иметь

$$\frac{d^2 \theta_1}{d\xi^2} - \xi \frac{d\theta_1}{d\xi} = -1, \quad \left. \frac{d\theta_1}{d\xi} \right|_{\xi=-\infty} = 0, \quad \theta_1(\xi)|_{\xi=\xi_*} = 0. \quad (5)$$

Решение системы (5) дает те же результаты, что приведены в табл. 1, только значения ξ и ξ_* берутся с противоположными знаками, как показано в табл. 2.

Анализ результатов

Изложенная выше методика определения времени изменения значений годового стока была апробирована на 5 основных реках Беларуси. Использовались данные за период инструментальных наблюдений в следующих створах: р. Припять – г. Мозырь, р. Неман – г. Гродно, р. Западная Двина – г. Витебск, р. Днепр – г. Могилев, р. Березина – г. Бобруйск.

Рассмотрим в качестве примера расчет времени изменения значений годового стока р. Припять – г. Мозырь. В табл. 3 приведены значения основных статистических параметров временного ряда годовых расходов воды р. Припять – г. Мозырь.

Таблица 3. Основные статистические параметры временного ряда годовых расходов воды р. Припять – г. Мозырь

$Q_{cp}, \text{ м}^3/\text{с}$	$\sigma, \text{ м}^3/\text{с}$	C_v	$r(1)$
390	123	0,32	0,29

Коэффициент k определяется по формуле $k = -\ln r = -\ln 0,29 = 1,24$.

Пусть в начальный момент времени $t = 0$ сток равен $Q = 640 \text{ м}^3/\text{с}$, а фиксированное значение стока $Q_* = 200 \text{ м}^3/\text{с}$. Тогда ξ определяется как отклонение начального значения стока от среднегодового в долях C_v , т. е.

$$\xi = \frac{Q - Q_{cp}}{Q_{cp} \cdot C_v} = 2.$$

Аналогично определяется ξ_* :

$$\xi_* = \frac{Q_* - Q_{cp}}{Q_{cp} \cdot C_v} = -1,5.$$

По значениям ξ и ξ_* из табл. 1 находим значение $\theta_1 = 6,0$. Тогда период времени, в течение которого значение стока будет находиться в пределах $[Q_*, \infty)$, определяется как частное θ_1 и k :

$$T_1 = \frac{\theta_1}{k} = 4,9.$$

Значения размерного времени в зависимости от значений Q и Q_* представлены в табл. 4.

Рассчитаем значения размерного времени, необходимые для моделирования гидрологических рядов, для оставшихся четырех створов. В табл. 5 приведены значения основных статистических параметров временных рядов годовых расходов воды для четырех створов.

Значения размерного времени в зависимости от значений Q и Q_* для створов р. Неман – г. Гродно, р. Западная Двина – г. Витебск, р. Днепр – г. Могилев, р. Березина – г. Бобруйск представлены в табл. 6 – 9.

Методика моделирования искусственных гидрологических рядов

Моделирование искусственных гидрологических рядов для створа р. Припять – г. Мозырь практически неограниченной длительности (более 1000 лет) производится с использованием результатов табл. 4, а также с применением простой цепи Маркова [4].

Таблица 4. Значения времени изменения годового стока р. Припять – г. Мозырь, лет

Фиксированное значение стока, м ³ /с	Сток в начальный момент времени, м ³ /с										
	200	250	300	350	400	450	500	550	600	650	700
200	0	1,8	2,8	3,5	3,9	4,3	4,5	4,7	4,9	5,0	5,2
250	0,2	0	1,0	1,7	2,1	2,5	2,7	2,9	3,1	3,2	3,4
300	0,4	0,2	0	0,7	1,1	1,4	1,7	1,9	2,1	2,2	2,3
350	0,7	0,5	0,3	0	0,4	0,8	1,0	1,2	1,4	1,5	1,7
400	1,0	0,9	0,7	0,4	0	0,3	0,6	0,8	0,9	1,1	1,2
450	1,6	1,4	1,2	0,9	0,5	0	0,3	0,5	0,6	0,8	0,9
500	2,4	2,1	1,9	1,7	1,3	0,8	0	0,2	0,4	0,5	0,6
550	3,6	3,4	3,2	2,9	2,6	2,0	1,3	0	0,2	0,3	0,4
600	5,9	5,7	5,5	5,2	4,8	4,3	3,6	2,3	0	0,1	0,3
650	10,7	10,5	10,3	10,0	9,7	9,1	8,4	7,1	4,8	0	0,1
700	22,0	21,8	21,6	21,4	21,0	20,5	19,7	18,4	16,1	11,3	0

Таблица 5. Основные статистические параметры временных рядов годовых расходов воды створов для основных рек Беларуси

Река – Створ	Q_{cp} , м ³ /с	σ , м ³ /с	C_v	$r(1)$
Неман – г. Гродно	197	35,5	0,18	0,16
Западная Двина – г. Витебск	227	61,6	0,27	0,31
Днепр – г. Могилев	143	34,9	0,24	0,22
Березина – г. Бобруйск	119	22,9	0,19	0,05

Таблица 6. Значения времени изменения годового стока р. Неман – г. Гродно, лет

Фиксированное значение стока, м ³ /с	Сток в начальный момент времени, м ³ /с										
	100	120	140	160	180	200	220	240	260	280	300
100	0	16,9	21,7	23,5	24,3	24,7	25,0	25,3	25,4	25,5	25,6
120	0,1	0	4,7	6,5	7,3	7,8	8,1	8,3	8,4	8,5	8,7
140	0,3	0,1	0	1,7	2,6	3,0	3,3	3,5	3,7	3,8	3,9
160	0,4	0,3	0,2	0	0,8	1,2	1,6	1,8	1,9	2,1	2,2
180	0,6	0,5	0,4	0,2	0	0,5	0,8	1,0	1,1	1,3	1,4
200	1,0	0,9	0,7	0,6	0,3	0	0,3	0,5	0,7	0,8	0,9
220	1,5	1,4	1,3	1,1	0,9	0,5	0	0,2	0,4	0,5	0,6
240	2,5	2,4	2,3	2,1	1,9	1,5	1,0	0	0,2	0,3	0,4
260	4,8	4,7	4,6	4,4	4,2	3,8	3,3	2,3	0	0,1	0,2
280	11,5	11,4	11,3	11,1	10,9	10,6	10,0	9,0	6,7	0	0,1
300	38,0	37,9	37,8	37,6	37,4	37,0	36,5	35,5	33,2	26,5	0

Таблица 7. Значения времени изменения годового стока р. Западная Двина – г. Витебск, лет

Фиксированное значение стока, м ³ /с	Сток в начальный момент времени, м ³ /с										
	100	125	150	175	200	225	250	275	300	325	350
100	0	5,1	7,5	8,8	9,6	10,2	10,6	10,9	11,1	11,3	11,4
125	0,2	0	2,4	3,8	4,6	5,1	5,5	5,8	6,0	6,2	6,4
150	0,3	0,2	0	1,3	2,1	2,7	3,1	3,4	3,6	3,8	4,0
175	0,6	0,4	0,2	0	0,8	1,4	1,8	2,0	2,3	2,5	2,6
200	0,8	0,7	0,5	0,3	0	0,5	0,9	1,2	1,5	1,6	1,8
225	1,2	1,0	0,9	0,7	0,4	0	0,4	0,7	0,9	1,1	1,3
250	1,7	1,6	1,4	1,2	0,9	0,5	0	0,3	0,5	0,7	0,9
275	2,5	2,3	2,1	1,9	1,6	1,3	0,8	0	0,2	0,4	0,6
300	3,7	3,5	3,4	3,1	2,9	2,5	2,0	1,2	0	0,2	0,4
325	5,9	5,7	5,6	5,3	5,1	4,7	4,2	3,4	2,2	0	0,2
350	10,4	10,2	10,0	9,8	9,5	9,1	8,6	7,9	6,6	4,5	0

В начальный момент времени путем розыгрыша моделируется случайное значение обеспеченности, затем по заранее определенной для временного ряда теоретической кривой обеспеченности определяется значение расхода, которое наносится на временную шкалу. Второе смоделированное значение обеспеченности дает возможность определить следующее значение стока. Имея значения стока в начальный и конечный моменты времени, по табл. 4 находится ближайший целый интервал времени, через который на шкалу наносится второе значение стока. Далее берется следующее значение обеспеченности, по которому определяется значение стока, и интервал вре-

мени, наносимый на шкалу. Аналогичным образом моделирование продолжается до момента достижения окончания временной шкалы, после чего процесс начинается сначала, но при заполнении шкалы в случае совпадения значений абсцисс приоритет отдается значению, найденному на более ранних этапах моделирования. По первому свободному значению на временной шкале определяется значение стока в предыдущий момент времени, а по нему – значение расхода в текущий момент времени с использованием простой цепи Маркова [4]:

$$x_{i+1} = 1 + r(x_i - 1) + \Phi_{i+1} C_v \sqrt{1 - r^2}, \quad (6)$$

Таблица 8. Значения времени изменения годового стока р. Днепр – г. Могилев, лет

Фиксированное значение стока, м³/с	Сток в начальный момент времени, м³/с										
	80	95	110	125	140	155	170	185	200	215	230
80	0	2,6	3,9	4,6	5,1	5,4	5,7	5,9	6,0	6,2	6,3
95	0,1	0	1,3	2,0	2,5	2,8	3,1	3,3	3,4	3,6	3,7
110	0,3	0,2	0	0,7	1,2	1,5	1,8	2,0	2,1	2,3	2,4
125	0,5	0,4	0,2	0	0,5	0,8	1,0	1,2	1,4	1,5	1,6
140	0,8	0,7	0,5	0,3	0	0,3	0,6	0,8	0,9	1,0	1,1
155	1,2	1,1	0,9	0,7	1,4	0	0,2	0,4	0,6	0,7	0,8
170	1,8	1,7	1,5	1,3	1,0	0,6	0	0,2	0,3	0,5	0,6
185	2,9	2,7	2,6	2,3	2,0	1,6	1,0	0	0,1	0,3	0,4
200	4,8	4,7	4,5	4,3	4,0	3,6	3,0	1,9	0	0,1	0,2
215	9,0	8,9	8,7	8,5	8,2	7,8	7,2	6,2	4,2	0	0,1
230	20,0	19,9	19,7	19,5	19,2	18,8	18,2	17,2	15,2	11,0	0

Таблица 9. Значения времени изменения годового стока р. Березина – г. Бобруйск, лет

Фиксированное значение стока, м³/с	Сток в начальный момент времени, м³/с										
	80	90	100	110	120	130	140	150	160	170	180
80	0	1,1	1,7	2,0	2,2	2,4	2,5	2,6	2,7	2,7	2,8
90	0,1	0	0,6	0,9	1,1	1,3	1,4	1,5	1,6	1,6	1,7
100	0,2	0,1	0	0,3	0,6	0,7	0,8	0,9	1,0	1,0	1,1
110	0,3	0,2	0,1	0	0,2	0,4	0,5	0,6	0,7	0,7	0,8
120	0,5	0,4	0,3	0,2	0	0,2	0,3	0,4	0,4	0,5	0,5
130	0,7	0,6	0,5	0,4	0,2	0	0,1	0,2	0,3	0,3	0,4
140	1,1	1,0	0,9	0,8	0,6	0,4	0	0,1	0,2	0,2	0,3
150	1,7	1,6	1,5	1,4	1,3	1,0	0,6	0	0,1	0,1	0,2
160	3,0	2,9	2,8	2,7	2,5	2,3	1,9	1,3	0	0,1	0,1
170	6,0	5,9	5,8	5,7	5,5	5,3	4,9	4,3	3,0	0	0,1
180	14,4	14,3	14,2	14,1	13,9	13,7	13,4	12,7	11,4	8,4	0

Таблица 10. Значения статистических параметров исходного и смоделированных временных рядов годовых расходов воды р. Припять – г. Мозырь

Временной ряд	Статистические параметры							
	Q_{cp} , м³/с	σ , м³/с	C_v	$r(t-1)$	$r(t-4)$	$r(t-5)$	$r(t-10)$	$r(t-24)$
Исходный	390	123	0,32	0,290	0,105	0,223	0,100	0,175
Смоделированный с помощью простой цепи Маркова	388	117	0,30	0,243	-0,063	-0,014	-0,061	0,101
Смоделированный нелинейным методом	399	137	0,34	0,089	0,115	0,004	0,121	0,049

где x_{i+1} – значение модульного коэффициента годового объема стока в $(i+1)$ -м году; r – коэффициент автокорреляции; x_i – значение модульного коэффициента годового объема стока в i -м году; C_v – коэффициент вариации; значение функции Φ определяется следующей формулой:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{z^2}{2}} dz.$$

Полученное значение годового стока в текущий момент является исходным для последующего моделирования. Процесс продолжается до тех пор, пока не будет заполнена вся шкала. Количество этапов ограничено и не может превысить длительность моделируемого гидрологического ряда.

Моделирование рядов годового стока указанным методом дает приемлемые результаты. Смоделированные ряды годовых расходов воды с применением простой цепи Маркова и нелинейным методом для створа р. Припять – г. Мозырь обладают статистическими параметрами, отличными от параметров исходного ряда в пределах $\pm 5-10\%$, как показано в табл. 10.

При помощи аппарата регрессионно-корреляционного анализа была получена сложная модель Маркова со сдвигом до 50

лет годовых колебаний расходов воды р. Припять – г. Мозырь, зависящая от $r(t-1)$, $r(t-4)$, $r(t-5)$, $r(t-10)$ и $r(t-24)$. Как показали исследования, моделирование искусственного гидрологического ряда простой цепью Маркова дает хорошие результаты только для автокорреляционной функции со сдвигом на один год, т. к. это параметр заложен при моделировании. Нелинейная модель позволяет прогнозировать ряд, обладающий схожей с исходным рядом корреляционной функцией со сдвигом на 4 и более года. Значения корреляционных функций для исходного и смоделированных рядов приведены в табл. 10.

Заключение

Показано применение стохастических дифференциальных уравнений для описания и прогнозирования многолетних колебаний годового стока. Решена задача стохастической гидрологии прогнозирования значения речного стока для 5 основных рек Беларуси, а также предложен метод моделирования искусственных гидрологических рядов, дающий лучшие результаты прогнозирования «дальней» корреляционной связи, нежели метод моделирования с использованием простой цепи Маркова. Результаты исследований можно применить при расчете и прогнозе многолетних колебаний речного стока неизученных и слабо изученных рек Беларуси.

СПИСОК ЦИТИРОВАННЫХ ИСТОЧНИКОВ

1. Найденов В.И., Швейкина В.И. Нелинейные модели колебаний речного стока // Водные ресурсы. Том 29, № 1. – М., 2002. – С. 62 – 67.
2. Пантелеев А.В., Якимова А.С., Босов А.В. Обыкновенные дифференциальные уравнения в примерах и задачах: Учеб. пособие. – М.: Высш. шк., 2001. – 376 с.
3. Прокопеня А.Н., Чичурин А.В. Применение системы Mathematica к решению обыкновенных дифференциальных уравнений: Учеб. пособие. – Мн.: БГУ, 1999. – 265 с.
4. Сванидзе Г.Г. Математическое моделирование гидрологических рядов. – Л.: Гидрометеиздат, 1977. – 296 с.

Статья поступила в редакцию 10.10.2005

УДК 004.8.032.6

Крапивин Ю.Б., Алькатауна Х.А., Страчук И.В.

ИСПОЛЬЗОВАНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПОИСКА ИНФОРМАЦИИ В СЕТИ INTERNET

Введение

Наиболее быстро развивающейся технологией является Интернет, которая объединяет множество разных сетей, миллионы компьютеров, около 300 миллионов пользователей со всех континентов, число которых по разным оценкам, увеличивается на 15-80% ежегодно.

Основная часть информации в Интернете - это неструктурированное хранилище информации огромного объема, которая характеризуется высокой динамичностью. В связи с чем, обеспечение поиска в Интернете становится критически важной задачей, которая не разрешима без соответствующих поисковых средств.

В статье рассматриваются современные поисковые системы в Интернете и основные технологии их построения, статистические закономерности естественно-языковых текстов, модели индексирования и поиска документов, а также использование нейронных сетей в поисковых системах.

Методы информационного поиска

Субъективно понимаемая цель идеального информационного поиска - найти все пертинентные и только пертинентные документы (найти «только то, что надо, и ничего больше»), при этом запрос должен максимально точно выражать информационную потребность, а документ должен быть максимально релевантным.

Информационный поиск производится при помощи систем информационного поиска как специальных комплексов программного, информационного и технического обеспечения.

Выполнение основных функций систем информационного поиска, обеспечивается её различными структурными элементами: информационно-поисковым языком, поисковой машиной, инструментами метапоиска, тематическими рубриками.

Как показывает проведенный анализ существующих поисковых инструментов в Интернете, все они имеют свои достоинства и недостатки.

По виду выдаваемой информации системы информационного поиска делят на документальные и фактографические. К документальным системам относятся поисковые машины, средства метапоиска, рубрикатеры.

Поисковая машина представляет собой, с одной стороны Web-сервер, главная страница которого обеспечивает пользователю возможность формирования запроса. С другой стороны, она обеспечивает создание и ведение каталога Web-страниц, который позволяет выбрать адреса нужных страниц по данным, содержащимся в запросе.

Схема, поясняющая организацию работы типичной поисковой машины, представлена на рисунке 1.

Каждая из основных универсальных поисковых машин покрывает ограниченное Web-пространство Интернет. По

различным оценкам, покрытие не превышает 30-40% доступных Web-страниц. При этом языковые возможности для записи поискового выражения также ограничены. Они не выходят за пределы ключевых слов и фраз, связанных операторами Буля (AND, OR, NOT) [1,2].

Метапоисковые средства позволяют расширить область поиска практически на всё Web-пространство Интернет, используя одновременно 6-12 основных универсальных поисковых машин. Однако выразительные средства языка формирования поискового выражения остаются теми же [3].

Наиболее совершенными поисковыми инструментами на сегодняшний день являются поисковые утилиты, так как они позволяют получить результаты поиска непосредственно на компьютер пользователя и самому пользователю выполнять их дополнительный анализ в режиме off-line. При этом возможно применение более мощного языка формирования поискового выражения [4].

Еще одним средством поиска информации в сети Интернет являются иерархические классификаторы (директории). Они обеспечивают рубрикацию по заданным тематикам как целых сайтов, так и других электронных ресурсов.

Классификацию текстов на естественном языке называют рубрицированием [5].

В настоящее время практическое применение получили следующие группы классификаторов:

- статистические классификаторы, на основе вероятностных методов.
- классификаторы, основанные на функциях подобия.
- классификаторы, использующие методы на основе искусственных нейронных сетей.

Классификаторы, использующие методы на основе искусственных нейронных сетей хорошо зарекомендовали себя в задачах распознавания изображений, и в данной статье рассматривается возможность их использования в обработке текстов на естественном языке.

Общие принципы функционирования системы

Предлагаемую поисковую систему можно логически разделить на следующие четыре части:

1. Интерфейс пользователя;
2. Индексирующий агент;
3. Модули системы;
4. База данных.

Укрупнённое взаимодействие этих частей показано на рисунке 2.

В основе функционирования описываемой системы лежит нейросетевой аппарат, обеспечивающий поиск и рубрикацию проиндексированных документов. Описания классов документов, представляют собой векторы действительных чисел, заложенные в синаптических весах искусственных нейронов, а

Крапивин Ю.Б., магистрант кафедры интеллектуальных информационных технологий БрГТУ.

Алькатауна Х.А., аспирант БГУИР.

Страчук Игорь Васильевич, ассистент кафедры технической эксплуатации автомобилей БрГТУ.

Беларусь, Брестский государственный технический университет, 224017, г. Брест, ул. Московская 267.