

УДК:

ПРЕДСТАВЛЕНИЕ ПОЛНОТЫ ИНФОРМАЦИИ, ИМЕЮЩЕЙСЯ В НАБОРАХ ДАННЫХ ПО АЭРОЛОГИЧЕСКИМ СТАНЦИЯМ

Арсений Тмофеев; Лидия Бадашова
ВНИИГМИ – МЦД, 6, ул. Королева., Обнинск, Калужская обл.,
249035, Россия
arsenit@yandex.ru

Целью настоящей работы являлась разработка технологии представления информации о полноте данных, содержащихся в глобальных радиозондовых архивах.

Разработанная технология позволяет строить наборы WEB страниц, содержащих таблицы и диаграммы, отражающие полноту и качество информации, представленной в соответствующем наборе. Особенности получаемого интерфейса являются: компактность, простота в использовании и наличие достаточного количества информации, необходимой для оценки полноты данных того или иного набора.

Введение

В настоящее время доступны, несколько глобальных наборов данных по аэрологическим наблюдениям (IGRA, CARDS, AEROSTAB и другие), которые активно используются в различных метеорологических исследованиях.

Имеющаяся сопроводительная информация (метаданные и документация) для всех таких наборов содержит чаще всего только описание формата записи данных, списки станций, включенных в наборы и периоды измерений, за которые представлены данные.

Однако, ряды измерений в наборах неоднородны и в процессе получения данных изменяются временное и пространственное (по количеству уровней) разрешения. Часть станций прекращали измерения на значительные сроки. Это приводит к тому, что для того, чтобы ознакомиться с набором и оценить полноту и качество представленных измерений, приходится зачастую выкачивать весь набор целиком и проводить самостоятельно небольшое исследование.

В то же время, для первоначального ознакомления с доступными данными, достаточно сравнительно небольшой информации о наборе. Это делает актуальной задачу обеспечения удобного интерфейса представления данных о полноте и содержанию наборов аэрологических данных.

Постановка задачи

В результате исследования структуры представленных для свободного доступа наборов, а также на основании собственного опыта проведения исследований на основе некоторых глобальных наборов к такому интерфейсу были сформированы следующие требования:

1. Ресурс должен обновляться регулярно.

Данные появляются в свободном доступе достаточно регулярно – в большинстве случаев – ежемесячно по мере накопления и обработки. Интерфейс должен предоставлять оперативную информацию – обновляться с минимально возможными задержками.

2. Доступность ресурса (небольшой объем).

Часто полоса пропускания, или доступный для использования Интернет трафик ограничены. Поэтому небольшой объем данных, представляющих собой интерфейс должен иметь максимально ограниченный размер.

3. Полнота представленной информации

Информации, представленной в интерфейсе должно быть достаточно для оценки полноты и качества представленной информации. В то же время излишние элементы могут привести к невыполнению предыдущего требования, следовательно, подбор элементов интерфейса должен быть проведен особенно тщательно.

4. Простота в использовании.

Ресурс должен быть наглядным, и должен позволять быстро получать необходимую информацию – то есть, пользователь, обладающий лишь поверхностными знаниями об аэрологических наборах без дополнительных инструкций должен иметь возможность сразу получить необходимую информацию

Технология представления данных

В рамках данной работы была разработана технология построения интерфейса, удовлетворяющего данным требованиям и представляющего собой набор HTML страниц.

Разработанная технология позволяет строить HTML страницы, содержащие таблицы и диаграммы, отражающие полноту данных в наборах, и позволяющие проводить быстрый визуальный анализ полноты и качества данных представленных в тех, или иных наборах. Также генерируемые страницы содержат статистическую информацию – основные статистики, рассчитанные для некоторых параметров, содержащихся в наборах.

Технология предполагает вывод двух типов данных по месячным и годовым порциям радиозондовых данных в формате HTML, готовом к использованию на веб-сервере.

Информация первого типа основана на анализе заголовков записей, относящихся к каждому отдельному запуску радиозондов и данных уровней, не включающей собственно значений метеорологических величин.

Информация доступна в виде таблиц и рисунков – диаграмм. Представлены такие параметры как:

- Общее количество запусков в месяц;
- Общее количество наблюдений за каждый день месяца;
- Общее количество наблюдений на уровне (для стандартных уровней давления);
- Статистики по качеству данных (Общее количество наблюдений с каждым значением флагом качества);
- Другие параметры;

В результате, информация о полноте данных для каждой метеорологической станции представлена 710 таблицами и рисунками – диаграммами.

Статистическая информация включает:

- Средние значения основных параметров на стандартных уровнях давления;
- Средние значения высоты и давления на уровнях тропопаузы;
- Средние значения высоты и давления на нескольких изотермических уровнях;

- Другие параметры;

Пример таблицы с основными статистиками приведен на рисунке 2.

Расчет статистических обобщений по исходным наборам данных требует обработки очень больших объемов данных. Поэтому все страницы и графики строятся один раз, при получении новой порции данных (обычно – раз в месяц), и после этого могут быть загружены на веб-сервер в виде статических HTML страниц и картинок. Это позволяет сохранить как дисковое пространство (месячный набор страниц занимает 3-10 МБ, включая картинки), так и процессорное время сервера (расчет статистик и построение страниц производится один раз и не требует дополнительных расчетов при отображении).

Все наборы данных по радиозондовым наблюдениям содержат данные в схожих (с теми или иными отклонениями) форматах. Общим для всех них является наличие двух принципиально разных типов записей:

1. Заголовочная запись – одна на каждый запуск зонда. (содержит, время, запуска, количество уровней измерения для данного зондирования и прочую описательную информацию.
2. Поуровневые записи – содержат собственно, значения изменяемых параметров, флаги качества, номер уровня, уровень давления, вспомогательные признаки уровня (уровень земли, тропопаузы, и т.п.)

Для обобщения информации о количественных характеристик измерений достаточно использовать только заголовочные записи, в то время, как для представления информации, например, о флагах качества, необходимо обработать набор целиком – включая поуровневые записи.

Для построения статических страниц используется статистический пакет SAS. Был разработан набор скриптов на языке SAS-Script, которые осуществляют:

- Ввод данных из исходных наборов
- Первоначальную обработку (перевод значений физических Величин в систему Си)
- Расчет обобщений (статистик) на исходных данных
- Вывод результатов в HTML страницы и рисунки

Примеры страниц, формирующих веб-интерфейс приведены на рисунках 1-6.

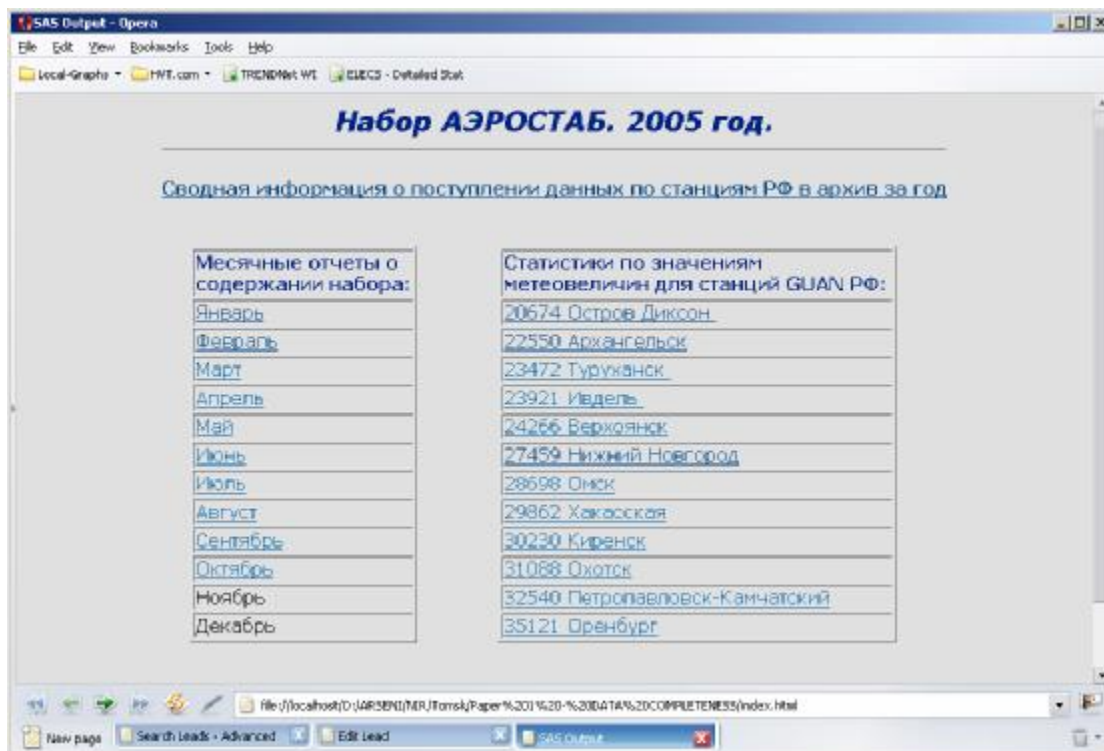


Рисунок 1 – Заголовочная страница

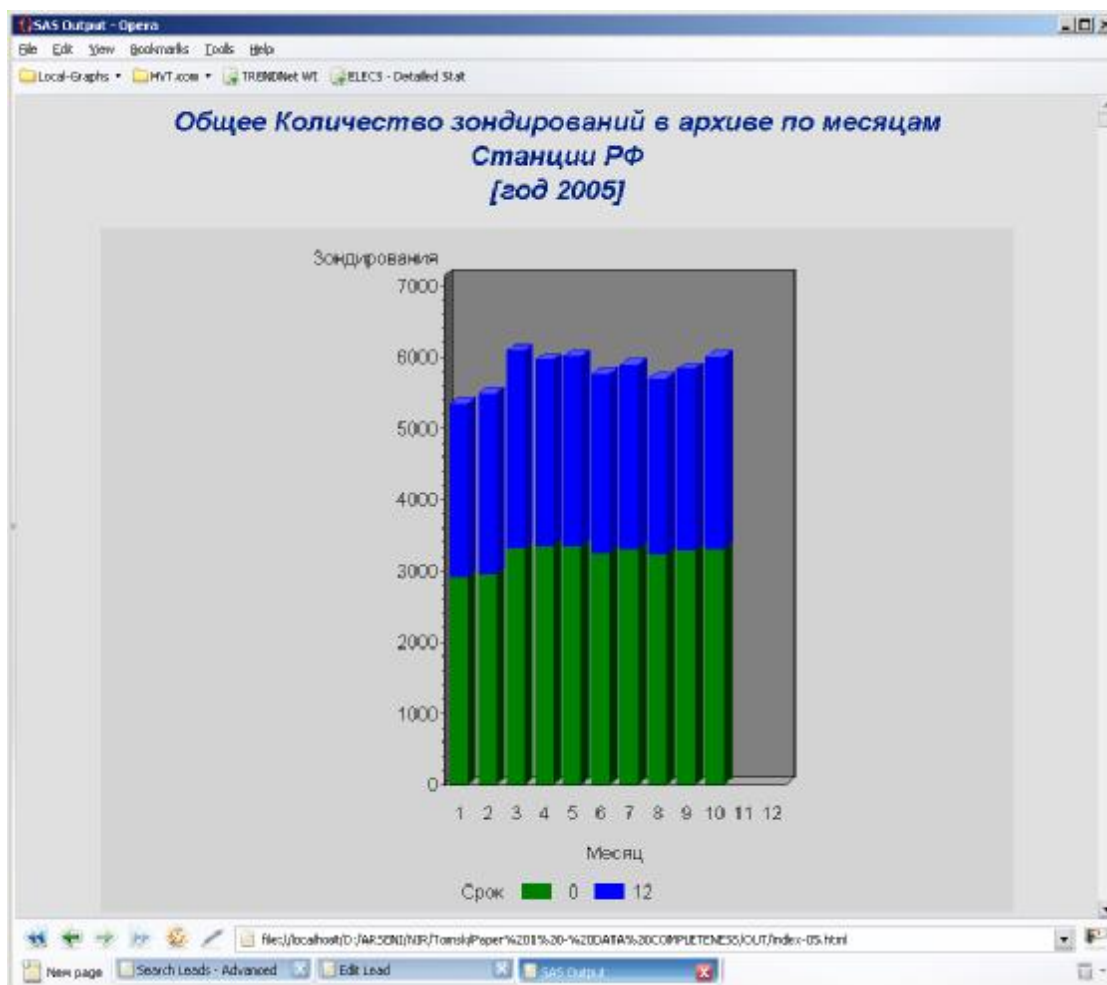


Рисунок 2 – Количество зондирований в наборе МОНАДС по месяцам 2005 года

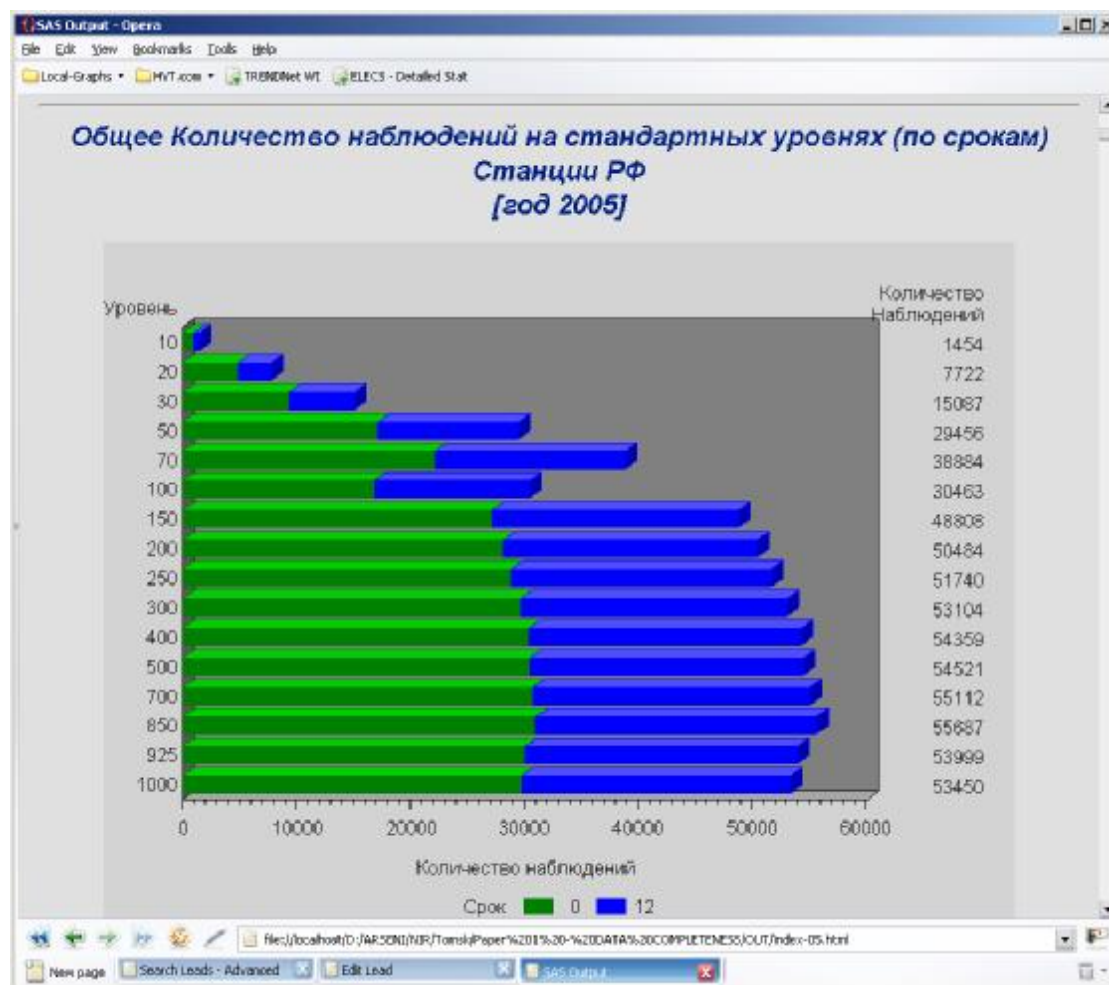


Рисунок 3 – Количество наблюдений на стандартных уровнях в наборе МОНАДС за 2005 год

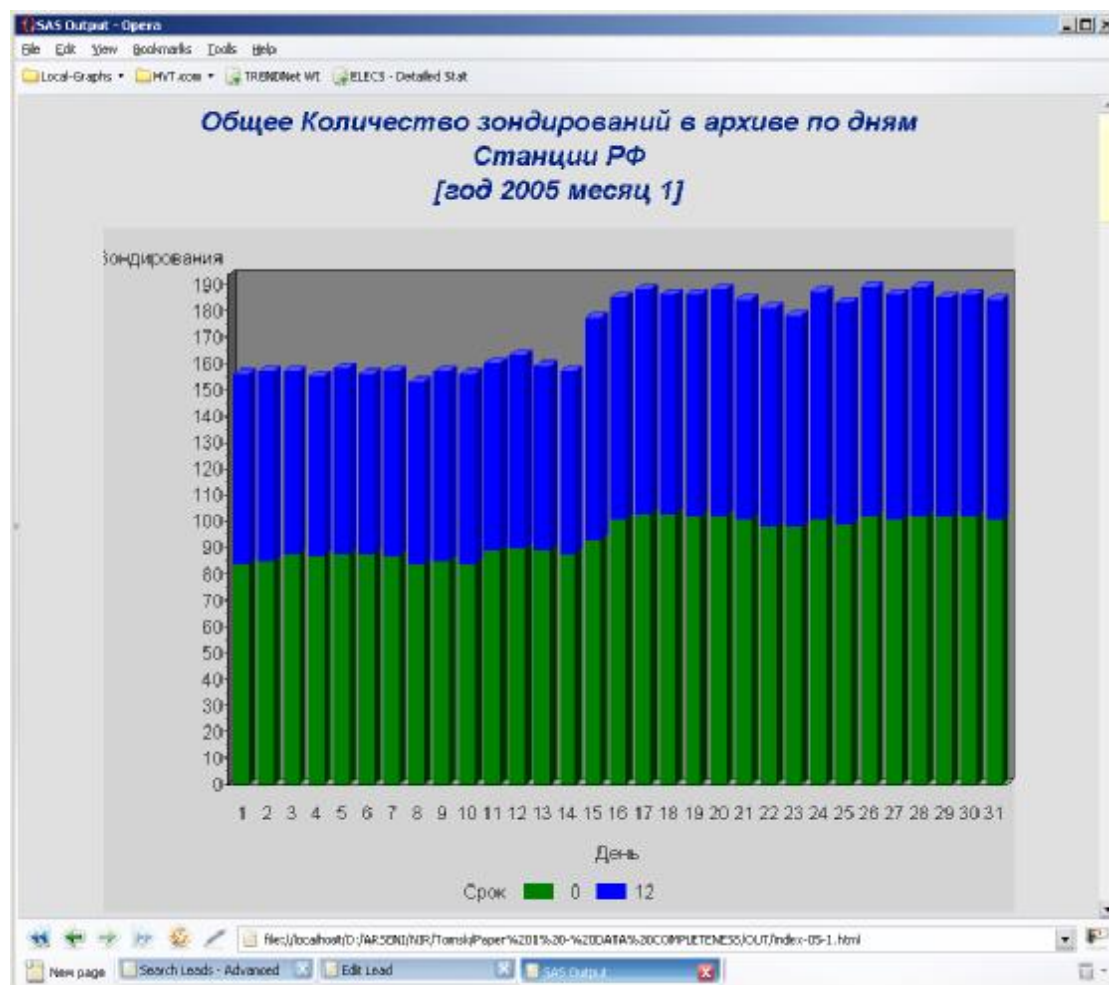


Рисунок 4 – Количество наблюдений на стандартных уровнях в наборе МОНАДС за 2005 год

SAS Output - Opera

File Edit View Bookmarks Tools Help

Local-Graphs MYT.com TRENDNet WL ELECS - Detailed Stat

**Количество зондирований в архиве по месяцам
Станции РФ
[год 2005]**

Количество зондирований	Месяц									
	1	2	3	4	5	6	7	8	9	10
Индекс станции										
20674	26	27	24	28	28	30	31	21	25	26
20744	21	20	21	23	20	22	27	23	26	25
20891	29	22	28	22	26	30	18	30	29	27
21824	30	28	27	59	60	59	62	62	60	59
21946	31	28	31	59	59	59	62	62	60	62
22113	62	55	61	58	60	59	50	62	60	62
22217	62	56	62	60	60	59	51	62	60	62
22271	27	25	28	29	19	28	56	50	35	53
22522	53	56	62	45	59	59	61	61	58	62
22550	62	55	61	47	57	59	62	62	51	61
22820	.	2	.	.	6	1	7	.	4	30
22845	62	56	62	49	57	58	58	56	59	52
23078	61	55	59	41	31	59	62	58	60	61
23205	62	56	62	60	52	59	62	61	52	62

file:///localhost/D:/AP-SONI/NIR/Tomsk/Paper%201%20-%20DATA%20COMPLETENESS/OUT/index-05.html

New page Search Leads - Advanced Edit Lead SAS Output

Рисунок 5 – Количество наблюдений в наборе МОНАДС за 2005 год

SAS Output - Opera

File Edit View Bookmarks Tools Help

Local-Graphs MYT.com TRENDNet WIT ELEC3 - Detailed Stat

Статистики по значениям метеовеличин
Станция 29862
2005 г. месяц 01

		M(T)	Std(T)	Q1(T)	Q3(T)	M(WS)	Std(WS)	Q1(WS)	Q3(WS)	M(P)	Std(P)	Q1(P)	Q3(P)	M(RH)	Std(RH)	Q1(RH)	Q3(RH)
Давление	Срок																
20	0	-49.90	8.24	-57.80	-41.50	26.00	2.83	24.00	28.00	26065.00	217.68	25950.00	26300.00	34.19	10.26	26.47	41.98
30	0	-54.35	7.35	-60.50	-47.90	31.80	8.87	24.00	41.00	23367.89	203.79	23240.00	23560.00	35.04	11.14	26.53	42.37
50	0	-54.36	6.02	-58.20	-49.90	27.46	8.23	20.00	36.00	20141.07	150.59	20055.00	20290.00	34.16	10.67	26.98	42.17
70	0	-55.40	4.68	-58.30	-52.10	23.65	6.30	18.00	29.00	17971.84	138.96	17830.00	18100.00	33.65	10.76	26.53	42.37
100	0	-55.82	3.21	-57.70	-52.90	19.40	6.24	14.00	26.00	15633.00	125.79	15540.00	15750.00	35.16	12.29	26.09	45.62
150	0	-55.50	3.45	-57.70	-53.30	16.45	5.96	11.00	20.00	13114.62	116.07	13000.00	13210.00	44.67	13.95	33.30	64.79
200	0	-57.35	5.41	-60.60	-53.60	16.68	7.61	13.60	20.00	11299.29	124.09	11190.00	11395.00	53.48	15.81	41.70	64.33
250	0	-58.51	4.83	-62.70	-55.10	18.13	9.24	13.00	24.00	9876.00	147.45	9740.00	10010.00	59.52	16.12	47.94	70.69
300	0	-56.18	2.52	-57.30	-55.10	19.48	9.66	13.00	26.00	8714.48	147.88	8580.00	8840.00	61.94	16.08	66.75	73.61
400	0	-46.11	4.66	-51.10	-42.30	18.73	8.30	16.00	22.00	6890.33	136.22	6730.00	6990.00	64.90	17.89	63.36	75.74
500	0	-36.51	5.69	-39.70	-31.30	16.84	7.51	12.00	19.00	5345.48	104.88	5250.00	5440.00	64.40	20.78	68.99	77.53
700	0	-21.53	5.80	-27.10	-16.10	10.33	5.66	8.00	12.00	2927.97	54.71	2878.00	2977.00	65.97	23.44	64.06	80.77
850	0	-14.44	6.98	-22.50	-10.30	5.39	3.69	2.00	7.00	1476.87	37.64	1461.00	1506.00	66.89	19.28	66.77	81.11
925	0	-13.97	6.76	-20.30	-8.10	4.65	2.90	2.00	6.00	834.23	41.76	815.00	859.00	74.77	7.89	69.73	78.91
1000	0	-27.73	6.40	-31.30	-22.50	0.38	0.66	0.00	1.00	250.07	52.38	219.50	292.00	73.10	8.13	67.44	77.59

file:///localhost/D:/AP/SONI/NIR/Tomsk/Paper%201%20-%20DATA%20COMPLETCNE35/OUT/index-v2005-str29862.html

New page Search Leads - Advanced Edit Lead SAS Output

Рисунок 6 – Статистики основных параметров свободной атмосферы

Использование SAS позволяет быстро строить и обновлять интерфейс, но используемые в этом случае статические страницы несколько усложняют организацию удобной навигации и поиска ресурсов.

Также, возможности подсистемы ODS (Output Delivery System), ограничивают возможности управления внешним видом страниц. Использование стилей и прочих дополнительных элементов возможно, но усложнено.

Это делает целесообразной некоторую доработку технологии. Одним из рассматриваемых путей совершенствования, является вынос последнего этапа построения интерфейса на веб-сервер. В этом случае все обобщения будут как и в настоящем варианте рассчитываться вне сервера, но непосредственно формирование HTML страниц будет происходить на сервере с использованием PHP/MySQL или аналогичных технологий. Это позволит использовать развитую систему поиска и отбора данных, а также позволит получить полный контроль над представлением данных.

Выводы

В результате выполнения работы, была разработана технология представления информации о полноте данных, содержащихся в глобальных архивах радиозондовых данных. Интерфейс удовлетворяет следующим требованиям:

- Регулярные обновления
- Компактный объем

- Полнота представленной информации
- Простота в использовании

Литература

1. В. Э. Иванов, М. Б. Фридзон, С. П. Ессяк Радиозондирование атмосферы. Технические и метрологические аспекты разработки и применения радиозондовых измерительных средств. Екатеринбург, 2004. 596 с.
2. Зверев А. С. Синоптическая метеорология. Л.: Гидрометеиздат, 1984.
3. Матвеев Л. Т. Курс общей метеорологии. Физика атмосферы. Л.: Гидрометеиздат, 1984.
4. DATA DOCUMENTATION FOR DATA SET 6300. National Climatic Data Center December 10, 2002
5. DESCRIPTION OF CONTENTS AND STRUCTURE OF MONADS National Climatic Data Center, May 20, 1998
6. Integrated Global Radiosonde Archive (IGRA) Format Description of IGRA. National Climatic Data Center, May 14, 2004

AEROLOGICAL DATASETS DESCRIPTIVE INFORMATION PRESENTATION TECHNOLOGY

Arseni Timofeev; Lidiya Badashova
 RIHMI – WDC, 6, Korolyov str., Obninsk, Kaluga Region,
 249035, Russia
arsenit@yandex.ru