

# РАЗВИТИЕ ЭКСПЕРТНОЙ СИСТЕМЫ «ОЗЕРА КАРЕЛИИ»

А. Л. Чухарев

*Институт водных проблем Севера КарНЦ РАН*

## Введение

Экспертная система «Озера Карелии», созданная В. В. Меншуткиным (Институт экономико-математических исследований РАН, СПб) совместно с сотрудниками Института водных проблем Севера [4, 5], является инструментом оценки неизвестных характеристик озер по имеющимся данным и эмпирическим зависимостям, установленным ранее. Вторая функция системы – выявление озер по заранее заданным пользователем параметрам.

Первый прототип системы был создан с помощью высокоуровневого языка программирования Visual Basic 6.0 для операционной системы Windows. К сожалению, на текущий момент времени этот язык программирования уже не поддерживается фирмой-создателем (Microsoft). Вследствие этого в качестве одной из целей представляемой работы было обозначено портирование приложения на более новую и современную программную платформу.

Первый прототип системы использовал базу данных по озерам Карелии, ограниченную 24 озерами и 22 параметрами (гидрофизических, химических и биологических характеристик) озер. В последующих версиях база данных была расширена до 100 озер. В базе данных параметры озера могли принимать только три значения: «низкое», «среднее», «высокое». Такой принцип соответствует нечеткой логике [1], когда система оперирует не с точными значениями физических величин, а с их приблизительными оценками. В настоящее время существует необходимость использования расширенных баз данных и распространения данного подхода при обработке больших массивов данных. Например, в ИВПС КарНЦ РАН создана база данных по 225 наиболее изученным озерам Карелии [6].

В числе целей данной работы можно назвать:

1. Создать веб-версию приложения для внедрения ЭС и использования ее широким кругом специалистов для практических и научных целей. Веб-интерфейс для новой версии приложения был выбран вследствие более удобного обновления и сопровождения экспертной системы. Также учитывался тот факт, что при работе с экспертной системой посредством технологии WEB пользователю не требуется устанавливать на компьютер дополнительное программное обеспечение. Используемые языки программирования: php, python. При портировании приложения сохранить всю функциональность и используемую логику работы с данными.
2. Использовать СУБД MySQL для хранения и пополнения данных, собранных сотрудниками Института водных проблем Севера.
3. Предусмотреть в ЭС опцию работы с базами данных, в которых присутствуют неизвестные значения параметров.
4. Создать гибкий способ добавления и расширения баз данных в ЭС.

## Материалы и методы

При создании программного продукта работа осуществлялась только с информационными материалами. Единственным необходимым оборудованием являлся компьютер. В Институте водных проблем Севера был установлен специальный сервер, на котором осуществлялась разработка и тестирование ЭС. Технологии, которые обеспечивают его функционирование, сведены в табл. 1. Критериями выбора тех или иных средств были: первое – способность решить поставленную задачу, второе – открытость лицензии, т. е. возможность свободного использования.

В результате проведения исследования было разработано веб-приложение. Веб-сервер установлен на компьютере в Институте водных проблем Севера. В настоящее время доступ к нему возможен только из локальной сети Карельского научного центра РАН. Через Интернет данный сервис пока не доступен.

Экспертная система состоит из нескольких компонентов. Во-первых, это база данных, содержащая информацию об озерах Карелии. Вторым компонентом является вычислительный обработчик базы данных. В настоящее время большая часть программного кода написана на языке про-

граммирования Python. База данных создана на основе информации справочника «Озера Карелии». В базе данных содержатся различные (гидрологические, гидрохимические, морфологические, географические) характеристики озер. Среди параметров можно отметить: бассейн, географические координаты (условного центра озера), площадь водосбора, высота над уровнем моря, генезис котловины, наличие притоков, прозрачность, площадь озера, длина береговой линии, объем озера, длина озера, ширина озера, глубина озера, показатели водообмена, характеристики термического и ледового режимов, химический состав воды, а также сведения о видовом составе макрофитов, фитопланктона, зоопланктона, макрозообентоса, ихтиофауны. В электронном виде эти базы доступны в формате xls (Microsoft Excel), а также в формате, используемом СУБД MySQL.

Таблица 1

**Используемые технологии**

Категория	Название	Комментарий
Языки программирования	Python, php	Обе технологии распространяются под свободными лицензиями. Имеется возможность использовать как в коммерческих, так и в некоммерческих целях. Также следует отметить, что Python имеет развитую библиотеку функций для проведения научных исследований. Официальный сайт проекта Python: <a href="http://www.python.org">www.python.org</a>
СУБД	MySQL	Свободная лицензия. Официальный сайт проекта: <a href="http://www.mysql.com">www.mysql.com</a>
Веб-фреймворк	Django	Свободная лицензия. Официальный сайт проекта: <a href="http://www.djangoproject.com">www.djangoproject.com</a>
Веб-сервер	apache2	Наиболее распространенный на настоящий момент сервер. Свободная лицензия сайт
Операционная система	Debian GNU/Linux 6.0	Дистрибутив свободной системы Linux. Официальный сайт проекта <a href="http://www.debian.org">www.debian.org</a>

В исходном виде база данных содержит точные значения этих параметров (так называемая четкая логика). Однако для того чтобы экспертная система могла работать с этими параметрами, они должны быть приведены к нечеткому виду. Нечеткий вид (нечеткая логика) подразумевает замену точных значений некоторой величины на ее количественную оценку (в простейшем случае – «много» или «мало»).

Таблица 2

**Описание значений нечеткой логики применительно к экспертной системе «Озера Карелии»**

Значение	Комментарий
0	Информация о характеристике неизвестна (не была измерена)
1	Очень низкое значение
2	Низкое значение
3	Среднее значение
4	Высокое значение
5	Очень высокое

Нечеткое множество для этих баз данных было расширено до 6 значений по сравнению с 3 значениями первой версии экспертной системы. Каждому значению было присвоен числовой идентификатор (см. табл. 2).

Принципиально новым здесь является наличие в базе данных значения 0. Это позволяет использовать базы данных с недостающими (неизмеренными) параметрами. Для этого потребовалось некоторое усовершенствование логики работы ЭС. При попадании нуля ЭС не изменяет внутренний массив вероятностей и считает, что данная характеристика не связана с другими. Расширение же значений нечеткого множества до 5 не требует качественных изменений в вычислительном ядре экспертной системы.

Подобный подход (переход к нечеткой логике) позволяет применять к физическим величинам аппарат дискретной математики и вероятностные оценки для реконструкции недостающих данных. Естественно, у нечеткой логики есть и недостатки. К ним можно отнести невозможность обратного перехода (к точным значениям физических величин).

Степень зависимости между величинами оценивалась как значение условной вероятности для параметра А иметь уровень N1 при условии, что параметр В имеет уровень N2

*Формула 1. Условная вероятность.*

$$P(A = N1 | B = N2) = \frac{P(A = N1 \cap B = N2)}{P(B = N2)}$$

Далее, при анализе неизученных озер появляется возможность использовать имеющиеся условные вероятности для реконструкции неизвестных параметров озер по известным. Например, известно, что широта и долгота неизвестного озера равны 62° с. ш. и 31° в. д. соответственно. Если перевести эти значения в нечеткие множества, используемые в описываемой системе, то они будут

равны 2 и 2. По имеющимся условным вероятностям известно, что вероятность для глубины иметь определенное значение (при известных координатах озера) описывается следующими значениями (см. табл. 3).

Первая колонка в таблице содержит уровни нечеткой логики, характеризующие глубину озера. Причем 1 соответствует самой маленькой глубине, 5 – самой большой. Вторая колонка содержит условные вероятности того, что озеро будет иметь глубину  $Z$  при координате  $\phi$ , равной 2. Третья колонка содержит условные вероятности того, что озеро будет иметь глубину  $Z$  при координате  $\lambda$ , равной 2. И, наконец, последняя колонка содержит произведение предшествующих двух вероятностей (пересечение двух независимых условий).

Теперь следует нормировать итоговые вероятности на единицу, чтобы оценить распределение озер в данных широтах по глубине. Из этой нормировки мы можем оценить вероятность для озера иметь определенную глубину при вышеописанных условиях. Очевидно, что нам следует умножить каждую из вероятностей на такое число, чтобы их сумма была равна единице.

$$\sum_{i=1}^n P_i * X = 1$$

Отсюда получаем,

$$X = \frac{1}{0.0016 + 0.0032 + 0.0038 + 0.0060 + 0.0002} = 67.57$$

Теперь получим вероятности распределения озер по глубине, удовлетворяющие нашим условиям.

В качестве интерпретации результатов можно сказать следующее. Если мы рассматриваем неизвестное озеро с координатами 62 и 31, то наиболее вероятно (40%), что его глубина будет от 137,64 м до 181,62 м (такой интервал соответствует уровню 4 в нечеткой логике). Очень большая глубина – от 181,62 до 225,6 м – практически невозможна – вероятность встретить такую глубину равна 1,3%. Глубины от 5,7 до 49,68 м, а также от 49,68 до 93,66 м и от 93,66 до 137,64 м возможны, но довольно маловероятны (их вероятности колеблются от 10 до 25%). Описанный алгоритм используется в экспертной системе для реконструкции неизвестных характеристик озер. Конкретизируя условия, можно уточнять вероятности для неизвестных величин иметь определенный уровень.

Довольно интересным представляется имплементация в экспертной системе теоремы Байеса, позволяющей по имеющемуся факту события вычислить вероятность того, что оно было вызвано определенной причиной (переставить причину и следствие местами). Однако в рамках данной работы подобное исследование не проводилось.

### Результаты и обсуждение

По итогам проведенной работы можно сделать вывод, что технологии экспертных систем являются применимыми в области лимнологии. В частности, подход с использованием нечеткой логики позволяет проводить анализ зависимостей между параметрами озер, не прибегая к их точным значениям. Говоря о базах данных, используемых в предлагаемой работе, нужно отметить, что они требуют некоторой фильтрации. Например, вследствие того, что определенные физические величины или параметры в базе данных не были измерены для большинства озер, представляется нужным не учитывать их в работе ЭС. Такие параметры не уменьшают степень неопределенности. Наоборот требуется оптимизировать работу ЭС с более известными параметрами. В качестве критерия отбора величин служит процент озер, в которых данная величина не измерена. Критическим уровнем было принято значение 0.9. То есть, если для 90% озер в базе данных не измерена данная величина, то такая величина не будет в дальнейшем учитываться ЭС. Такой подход позволяет оптимизировать работу приложения.

Таблица 3

**Посчитанные, исходя из доступных значений базы данных, условные вероятности, описывающие глубину озера в зависимости от координат**

Z	P (Z при $\phi = 2$ )	P (Z при $\lambda = 2$ )	P (Z)
1	0,08	0,02	0,0016
2	0,16	0,02	0,0032
3	0,13	0,03	0,0038
4	0,12	0,05	0,0060
5	0,01	0,02	0,0002

Таблица 4

**Вероятности для озера иметь определенную глубину, исходя из заданных условий**

Z	P(Z)
1	0,1080
2	0,2160
3	0,2560
4	0,4050
5	0,0130

Большинство параметров, на основе которых построена экспертная система, будут опубликованы в виде справочника «Озера Карелии». В табл. 5 приведена сравнительная характеристика двух способов публикации данных: традиционного – в виде бумажного справочника и инновационного – в виде информационного сервиса экспертной системы.

Таблица 5

**Сравнительная характеристика справочника «Озера Карелии» и web-сервиса «Озера Карелии»**

Характеристика	Справочник на бумажном носителе	Web-сервис
По способу обновления	В справочник на бумажном носителе изменения вносить сложно. Требуется публикация нового издания книги, содержащего исправления и дополнения	Исправления вносятся быстро и становятся доступными с момента изменения базы данных
По способу доступа к данным	Доступ к информации осложнен отсутствием автоматизированных средств поиска и обработки информации. Например, для поиска всех озер, глубина которых больше 50 м, потребуется вручную просмотреть весь справочник	Быстрый доступ к информации. Существует возможность поиска в базе данных по любым параметрам
По способу масштабирования	Распространение информации ограничено тиражом справочника	Распространения информации ограничено сетью Интернет
Дополнительные свойства	Нет	Функциональность экспертной системы

Говоря о функционале экспертной системы, в числе главных ее возможностей следует отметить оценку неизвестных параметров озер по известным. В основе прогнозирования лежат эмпирические закономерности, выявленные в результате анализа больших массивов информации. Подобные методики дают вероятностный результат. Однако, увеличивая количество озер, по которым производится анализ, можно повышать достоверность оценки.

**Заключение**

В ходе проведения данной работы применены технологии экспертных систем. На примере конкретного озера были продемонстрированы результаты работы экспертной системы «Озера Карелии». Была показана работоспособность подхода использования нечеткой логики в предметной области, а также намечены возможные направления для дальнейших исследований.

В дальнейшем предполагается выполнить следующие усовершенствования ЭС для Веб-приложений:

1. Интеграция с геоинформационной системой (ГИС на платформе MapInfo). Система будет по запросу показывать на карте искомое озеро (озера или водосборы).
2. Создание все более «дружественного», простого для широкого круга специалистов пользовательского интерфейса.
3. Организация доступа к ЭС через Интернет.

**Литература**

1. Заде Л. А. Понятие лингвистической переменной и его применение к принятию решений. М., 1976. 165 с.
2. Меншуткин В. В. Искусство моделирования (экология, физиология, эволюция). Петрозаводск; СПб., 2010.
3. Меншуткин В. В., Китаев С. П., Филатов Н. Н., Чухарев А. В. Экспертная система «Озера Карелии».
3. Модельный подход к изучению сообществ рыб // Водные ресурсы. 2010 (в печати).
4. Меншуткин В. В., Филатов Н. Н., Потахин М. С. Разработка экспертной системы «Озера Карелии». Ч. 1. Порядковые и номинальные характеристики озер // Водные ресурсы. 2009. Т. 36, № 2. С. 160–171.
5. Меншуткин В. В., Филатов Н. Н., Потахин М. С. Разработка экспертной системы «Озера Карелии»: Ч. 2. Классификация озер // Водные ресурсы. 2009. Т. 36, № 3. С. 300–311.
6. Озера Карелии. Гидрология, гидрохимия, биота: Справочник / Под ред. чл.-корр. РАН Н. Н. Филатова и к. б. н. В. И. Кухарева. Петрозаводск (в печати).