Classifying living things with computer science

An activity by the Australian Computing Academy

Activity overview

What's it all about?

The activity uses a series of questions focusing on structural features of plants and animals to correctly identify the target species.

Through the careful selection of appropriate questions, students can determine which species has been selected in the minimum number of steps.

This is a fundamental concept of data classification in computer science, biology and many other academic fields.

It is accessible to students of all ages, and ties in nicely with both the Digital Technologies and Science subjects in the Australian Curriculum.

Resources needed

Preparing for the activity

The activity doesn’t require computers - everything you needed is included in this presentation and linked resources.

Students will need printed copies of the necessary slides, and you can guide them through the activity using the slides in the relevant section of this presentation.

You can turn this into an optional programming activity if students have access to computers and would like to take it further.

Adaptations and extension suggestions are included at the end of this presentation.

Introducing the topic

What do the students need to know before they start?

Before we begin, we'll frame the activity in terms of the need for classification of data sets, and introduce some terminology students should know.

The introduction and example activity in this slide deck uses physical characteristics of living things for the basis of classification, in line with the expectations of the Year 3 curriculum ACSSU044.

The same activity can be done with more complicated concepts. Blank templates for all stages of the activity are included in the extension/adaptation section at the end of the slide deck.

Classifying living things

What is classification?

Classification describes grouping things together based on common features. It allows us to identify living and non-living things that share similar characteristics. We can look for physical and behavioural properties that things have in common, and use this to communicate what we are talking about.

There are lots of different living things on Earth. Scientists use classification to understand the diversity of the many living things we share our planet with. It allows them to make conclusions about how different species live and interact with the world.

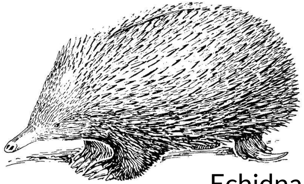
What is an algorithm?

An *algorithm* is a sequence of steps for solving a problem or completing a task. The steps can include questions that help us decide what our next steps are going to be. We call this *branching*.

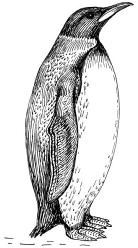
If we think carefully about the questions we ask when we classify plants and animals, we can design an algorithm that ensures we always end up identifying the correct species from our collection. This is what computer scientists do all the time when working with different types of data!

Example: Classifying living things

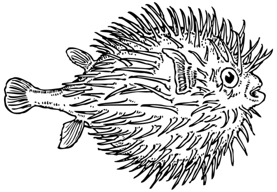
Take a look at the following species:



Echidna

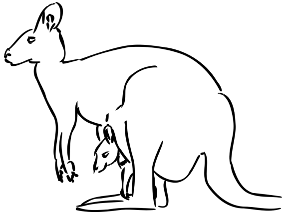


Penguin

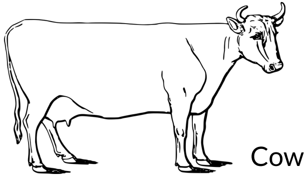


Puffer fish

Platypus

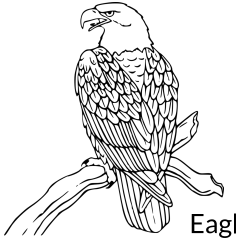


Kangaroo



Cow

Frog



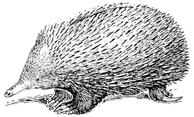
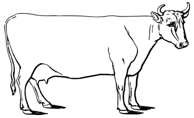
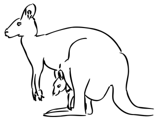
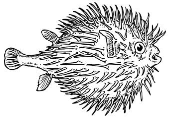
Eagle

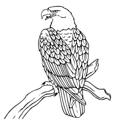
Identifying features

To classify things into groups, we need to look for features that are either common or distinctive between each one. To create our algorithm, we then need to work out how those features can be identified using simple yes/no questions.

What features do you see in the species shown?

You can use a table to summarise your observations, and this will help us build an algorithm to identify each one of our species.


Species	*Fur?*	*Feathers?*	*Spiky?*	*Fly?*	*Pouch?*	*Lay eggs?*
Echidna						
Kangaroo						
Penguin						
Cow						
Puffer fish						
Frog						
Platypus						
Eagle						

Species	*Fur?*	*Feathers?*	*Spiky?*	*Fly?*	*Pouch?*	*Lay eggs?*
Echidna	✔	×	✔	×	✔	✔
Kangaroo	✔	×	×	×	✔	×
Penguin	×	✔	×	×	×	✔
Cow	✔	×	×	×	×	×
Puffer fish	×	×	✔	×	×	✔
Frog	×	×	×	×	×	✔
Platypus	✔	×	×	×	×	✔
Eagle	×	✔	×	✔	×	✔

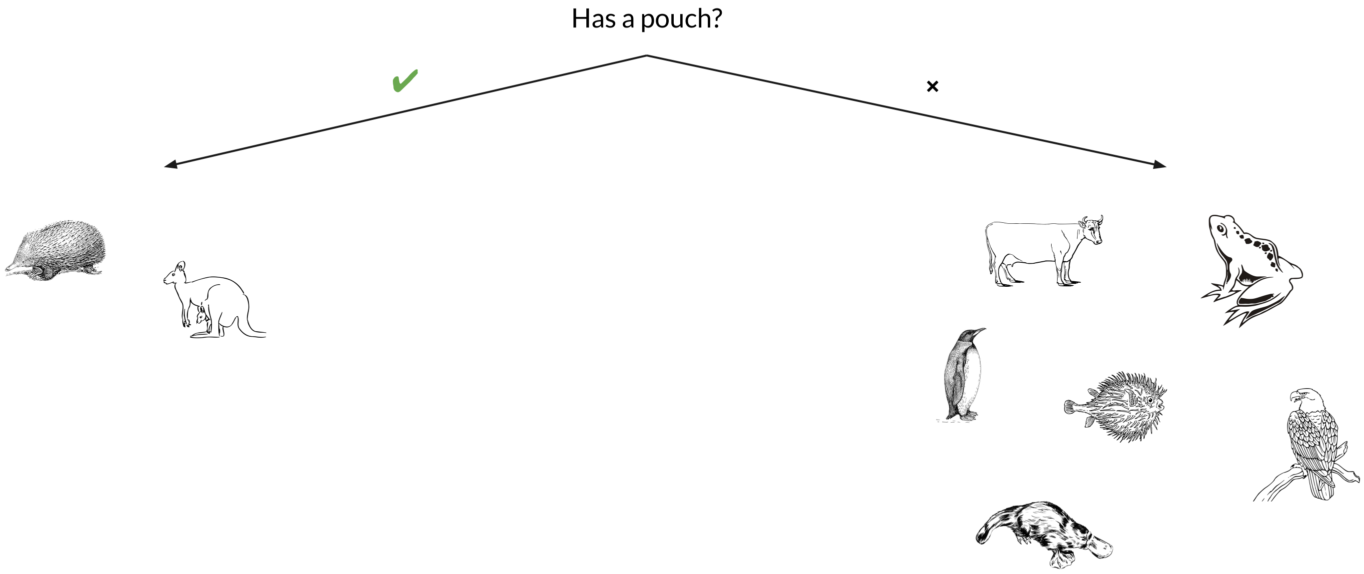
Some of the features of each species

Decision trees

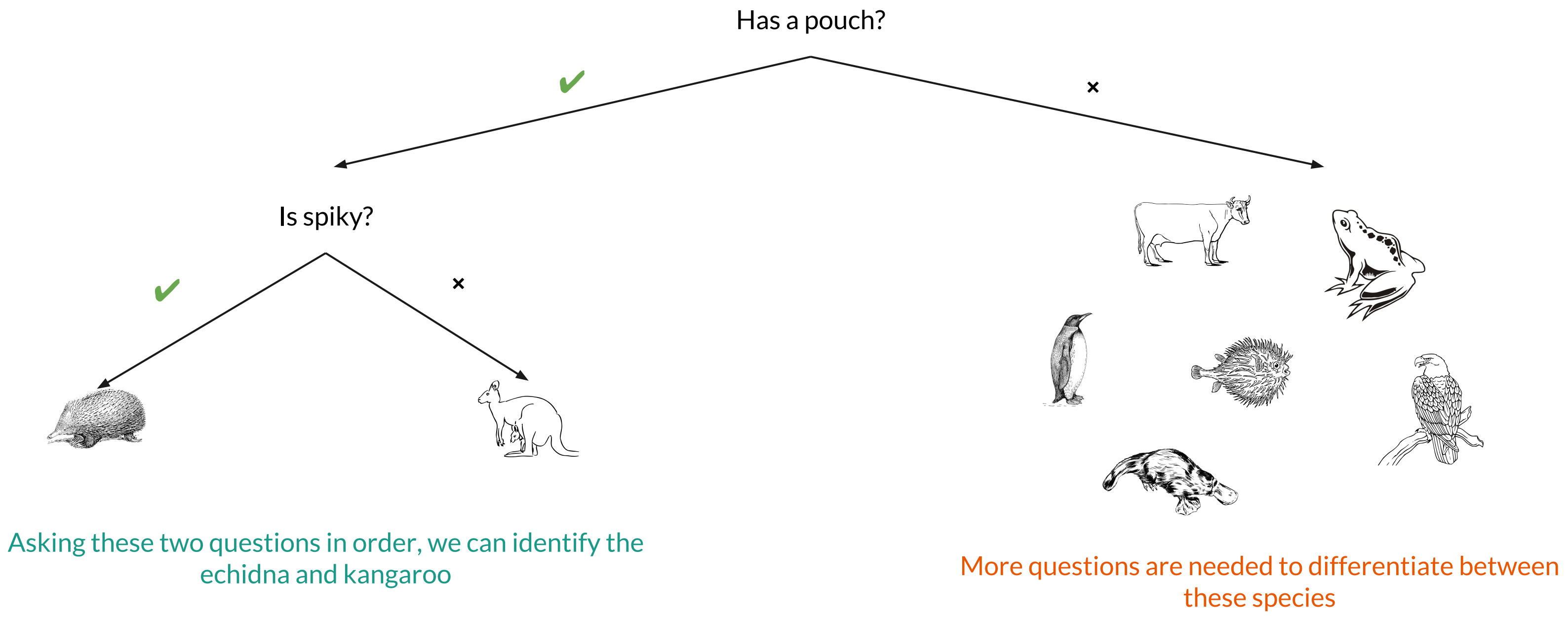
We can now create a *decision tree* that we can follow to correctly identify each of our species.

A decision tree is an ordered series of yes/no questions that allows us to narrow down our selection based on what we know about each species. We ask each of the questions in order as we move from the top to the bottom of the tree.

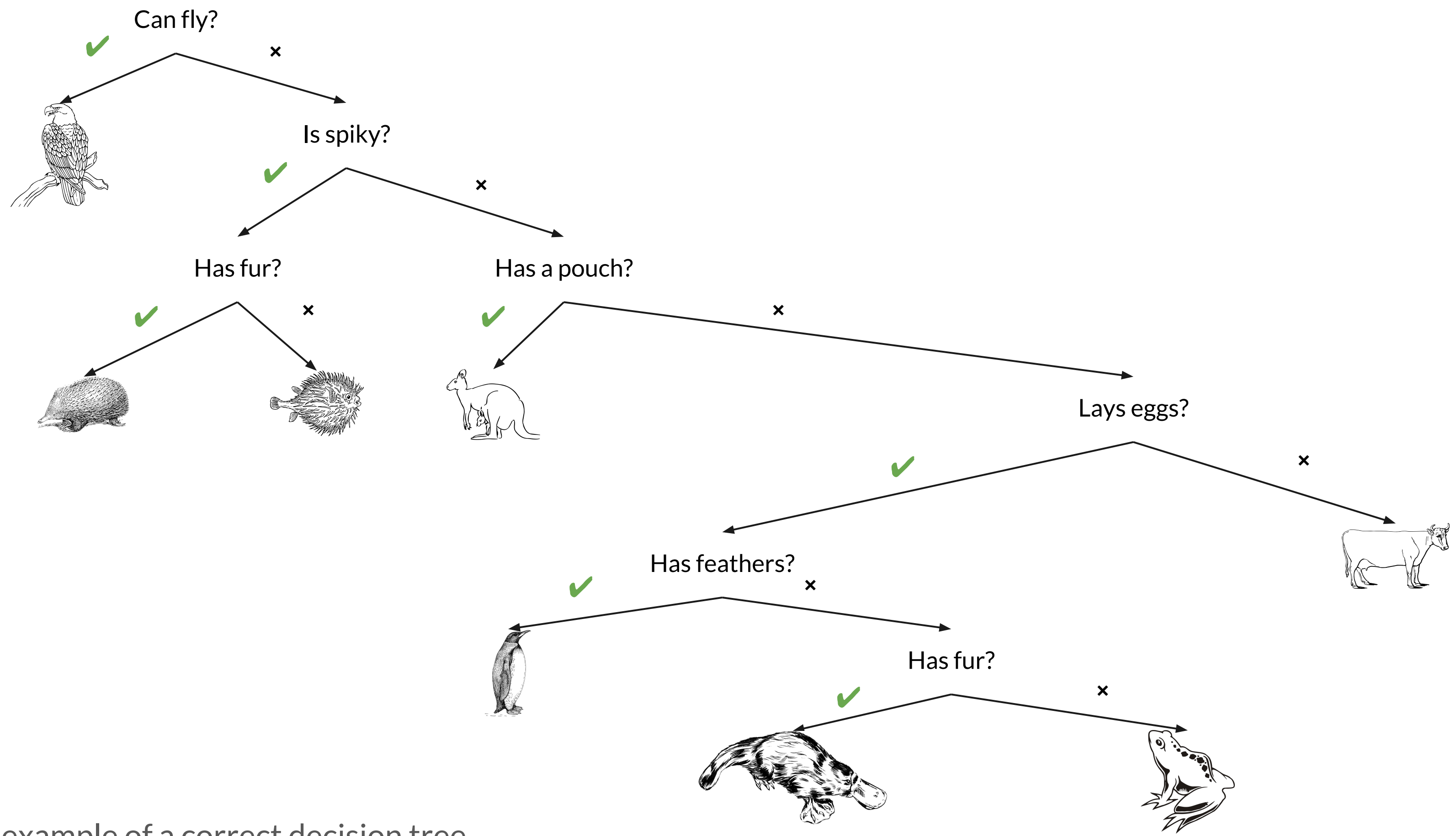
A decision tree is a visual representation of the algorithm we will be using each time we want to identify our selected species. It will work for all of the species we are using.



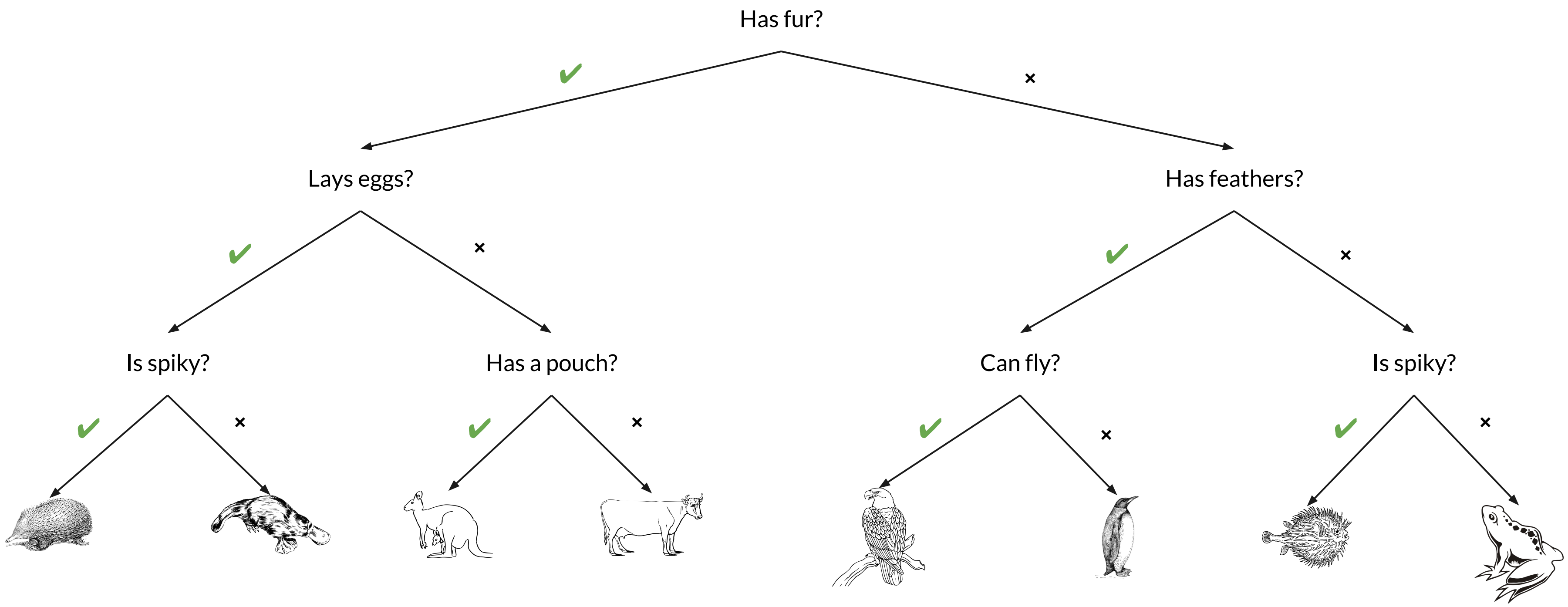
An example of one branching question



Splitting up the animals with a pouch



One example of a correct decision tree



Another example of a correct decision tree

Order matters

It is really important to think about the order that you are asking questions. You can ask the same question at different times in the tree, and end up with very different results.

Notice too how in the second example (the balanced tree) that you can ask “Is spiky?” at different times and get a different result? If you asked that question earlier, it may not allow you to successfully identify the puffer fish or the echidna, since there may be other features you haven’t yet eliminated in your classification.

For example, both the echidna and the puffer fish lay eggs, so if you asked only those two questions, then you wouldn’t be able to differentiate between them. You would still need to check if the animal had fur (or perhaps something else, like if the animal can breathe underwater).

Evaluating decision trees

Some decision trees are better than others. The best decision trees will try to split the group in half each time, since that will keep the number of questions that need to be asked to a minimum.

In our first tree, it only took one question to identify the eagle, since it is the only species that can fly. But, if we had been identifying the platypus or frog, we needed to ask six questions to know which one we had. Notice how the decision tree looks lopsided, or really long one one side?

In our second tree, we only ever have to ask three questions, no matter which animal has been chosen. The tree is balanced - it is about the same size no matter which branch you go down.

When designing an algorithm to identify our species, we should always try to make a balanced tree where possible, since this means we will always be able to classify our species relatively quickly.

Exploring things further

Doing the activity

Now that students have been shown an example activity that demonstrates how to classify things and generate a decision-tree, they should be given an opportunity to do the same with a larger set of objects.

The following slides frame a challenge for students, and then provide you with a few different collections of animals for classification. Each collection has been given a “difficulty” rating to help you choose an appropriate collection for your students.

You may choose to put constraints on the questions they can ask (e.g. for older students, the questions may have to be about the habitat of the species) once they understand how the activity works.

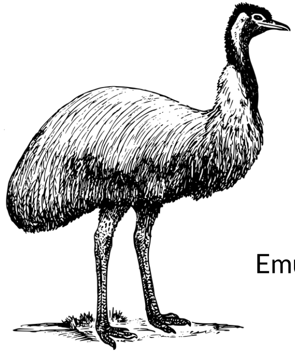
Activity: Designing a decision tree

What is the best way to classify these species?

The worksheets you've been given include images of different species. Using your knowledge of those species, you need to develop an algorithm that will allow you to identify each species efficiently, and represent that algorithm as a decision tree.

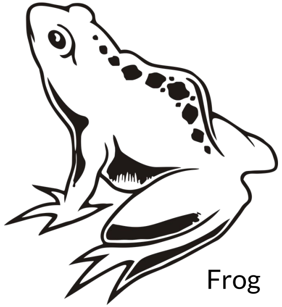
It will help to draw up a classification table first so that you can identify the features that will allow you to split the group evenly. Then, you can write out your questions and draw your decision tree.

You should test your tree out with another student to make sure that they agree that your questions and answers are correct. Who can come up with the best classification?

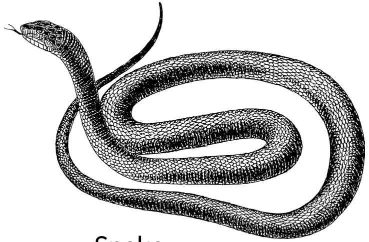


Emu

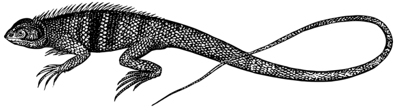
Kookaburra



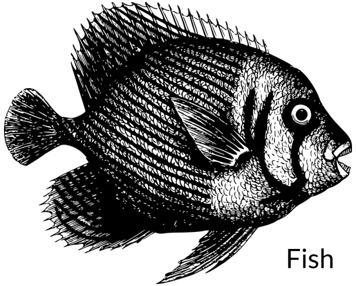
Frog



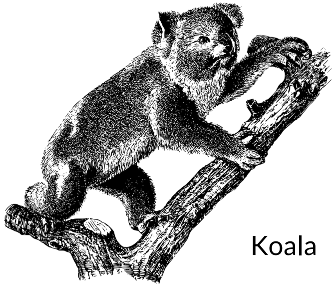
Snake



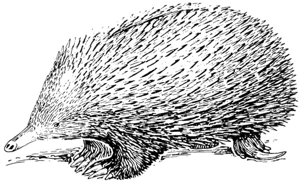
Lizard



Fish

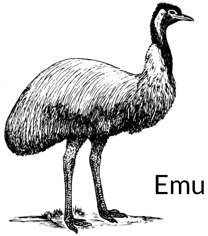


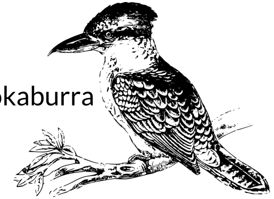
Koala



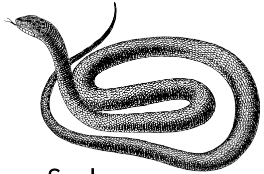
Echidna

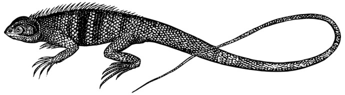
Standard collection of species

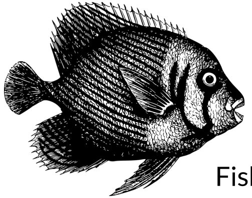

Emu

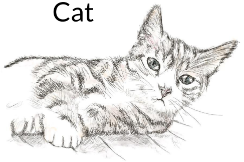

Kookaburra

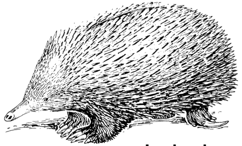

Frog

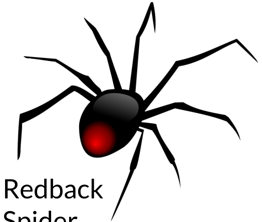

Snake

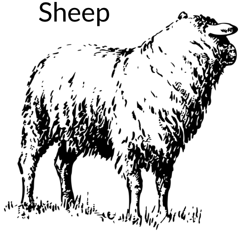

Lizard

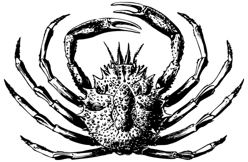

Fish

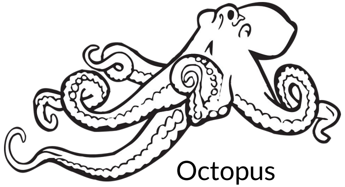

Cat

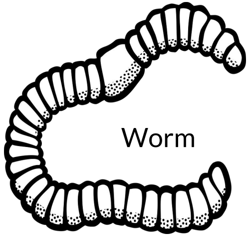

Echidna

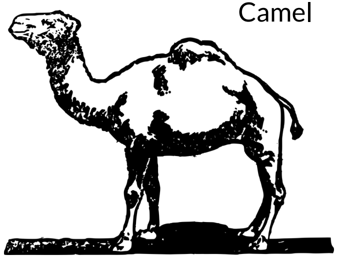

Redback Spider


Sheep

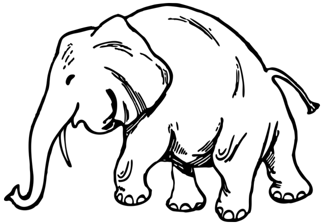

Crab


Octopus


Worm


Camel


Giraffe

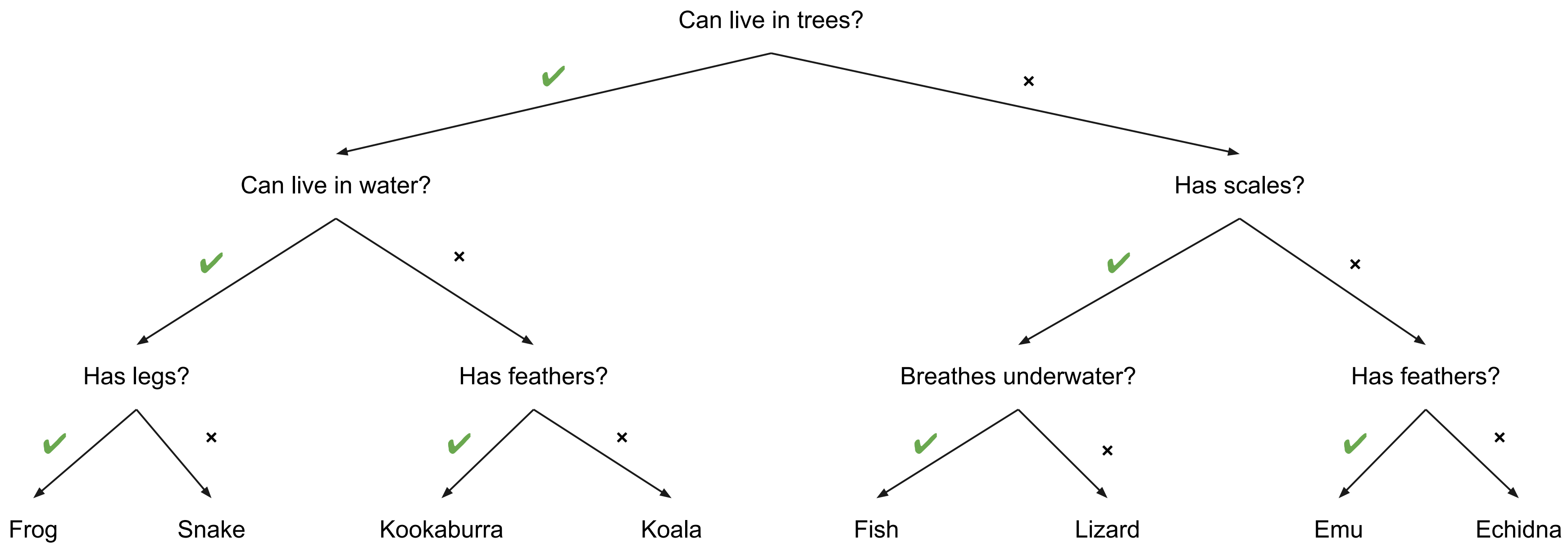

Elephant

Challenging collection

Teacher example answers

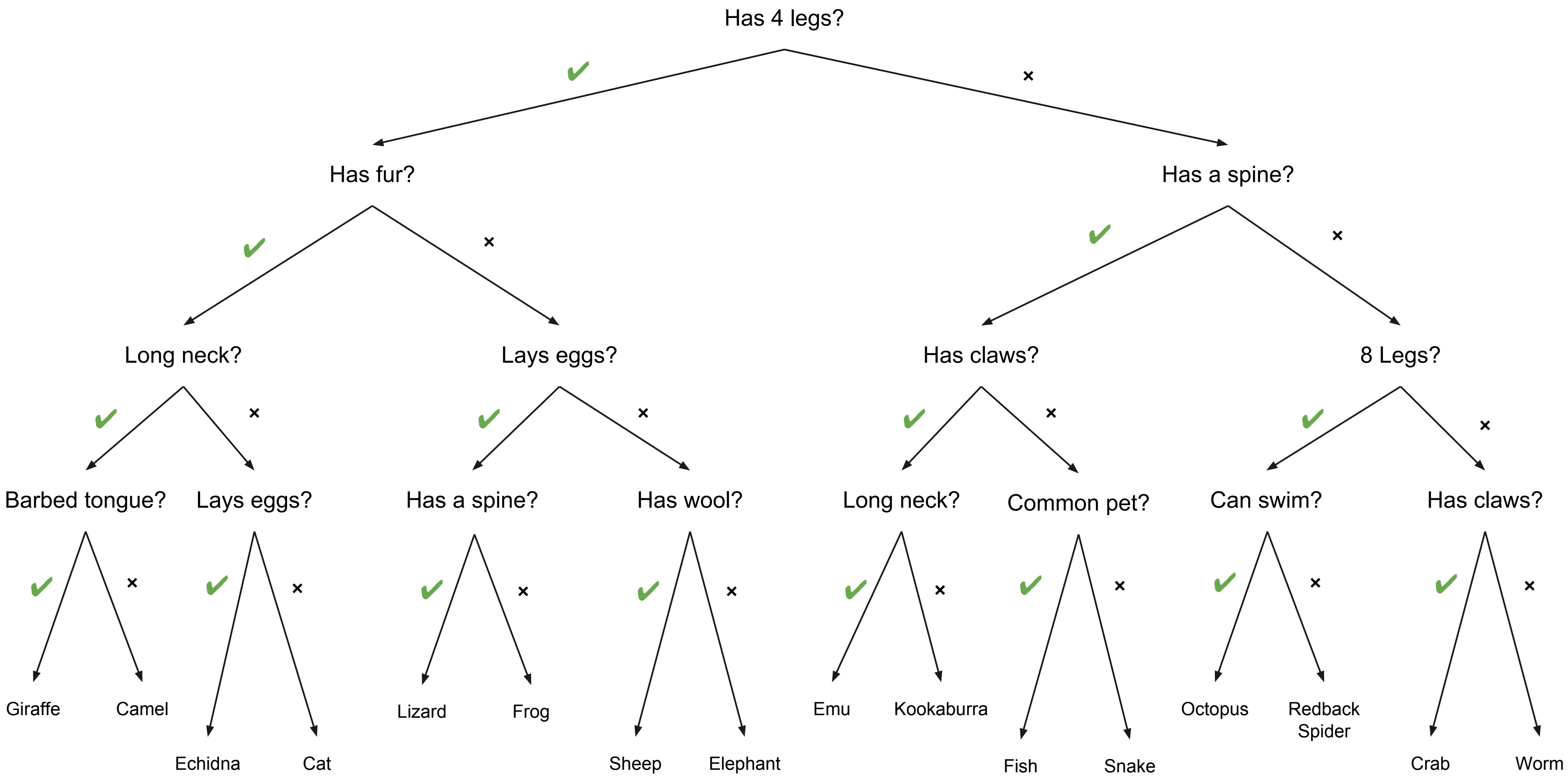
Species	*Scales?*	*Feathers?*	*Live in trees?*	*Legs?*	*Live in water?*	*Breathes underwater?*
Emu	×	✔	×	✔	×	×
Kookaburra	×	✔	✔	✔	×	×
Frog	×	×	✔	✔	✔	×
Snake	✔	×	✔	×	×	×
Lizard	✔	×	×	✔	×	×
Fish	✔	×	×	×	✔	✔
Koala	×	×	✔	✔	×	×
Echidna	×	×	×	✔	×	×

Sample classification table for standard collection of species



Sample balanced solution for the standard collection of species

Species	*4 Legs?*	*Has a spine?*	*8 Legs?*	*Can swim?*	*Has fur?*	*Lays eggs?*	*Long neck?*	*Has wool?*	*Has claws?*	*Common pet?*	*Barbed tongue?*
Emu	✕	✔	✕	✕	✕	✔	✔	✕	✔	✕	✕
Kookaburra	✕	✔	✕	✕	✕	✔	✕	✕	✔	✕	✕
Frog	✔	✕	✕	✔	✕	✔	✕	✕	✕	✕	✕
Snake	✕	✔	✕	✔	✕	✔	✕	✕	✕	✕	✕
Lizard	✔	✔	✕	✔	✕	✔	✕	✕	✕	✕	✕
Fish	✕	✔	✕	✔	✕	✔	✕	✕	✕	✔	✕
Cat	✔	✔	✕	✔	✔	✕	✕	✕	✔	✔	✔
Echidna	✔	✔	✕	✔	✔	✔	✕	✕	✔	✕	✕
Redback Spider	✕	✕	✔	✕	✕	✔	✕	✕	✕	✕	✕
Sheep	✔	✔	✕	✔	✕	✕	✕	✔	✕	✕	✕
Crab	✕	✕	✕	✔	✕	✔	✕	✕	✔	✕	✕
Octopus	✕	✕	✔	✔	✕	✔	✕	✕	✕	✕	✕
Worm	✕	✕	✕	✔	✕	✕	✕	✕	✕	✕	✕
Camel	✔	✔	✕	✔	✔	✕	✔	✕	✕	✕	✕
Elephant	✔	✔	✕	✔	✕	✕	✕	✕	✕	✕	✕
Giraffe	✔	✔	✕	✔ (very badly)	✔	✕	✔	✕	✕	✕	✔



Sample balanced solution for the challenging collection of species

Reflection and evaluation

What have we learned?

It is important that students get a chance to reflect on their learning, and to evaluate their own solutions alongside the solutions of other students.

You can begin by having students discuss the differences between theirs and a peer. You can also have them compare their solution to the sample one, and indicate in which circumstances theirs will perform faster or slower.

Having them think about how computers might use this approach in the apps or data sets they are familiar with is a good way of evaluating their understanding. For example, being able to quickly eliminate / identify data based on its properties is useful for searching and filtering.

Adaptation and extension activities

What else can we do with this idea?

If you're looking for ways to use this type of activity in different contexts, the following slides provide some suggestions that you can use to consolidate these ideas or extend students.

An indicative year level based on the standards expressed generally in the Australian Curriculum has been provided as a guide, but this isn't a hard and fast rule.

Other interesting things to classify

Classification activities become really interesting when there are a lot of different features or properties associated with each item in your collection. Other things that provide a lot of scope for interesting classification challenges (and are accessible to any age group) include:

- Plants
- Flags
- Artwork, movies or music
- Pokémon (some students really enjoy this!)

Framing each of your classification challenges in terms of a problem (e.g. If you're a traveller, what kind of grouping of flags makes the most sense? Politics? Sport?) gets students thinking about how the context of the problem drives decisions around classification. You can draw this back to the natural split of plants and animals in biological classification, and talk about how conventions are established.

Encoding the data

For high school students (curriculum bands 7-8 and 9-10), once the data has been classified and a decision-tree established, thinking about alternative representations of the same data provides scope to introduce other aspects of Information Theory. For example, if we know that our software will always ask the user a specific set of questions in a pre-defined order, then we can map the answers to binary values:

- *Yes, No, No, Yes* would be 1001
- *Yes, Yes, No, No* would be 1100

We can then use a lookup table to determine what output we will provide to the user. Quizzes such as “Which Harry Potter character are you?” on social media often use this kind of approach, where your answers are used in a lookup table to generate your quiz result. This can be turned into a programming activity for students to write that uses more complicated data structures, but simplifies the maintenance of the software.